Monte Carlo Integration

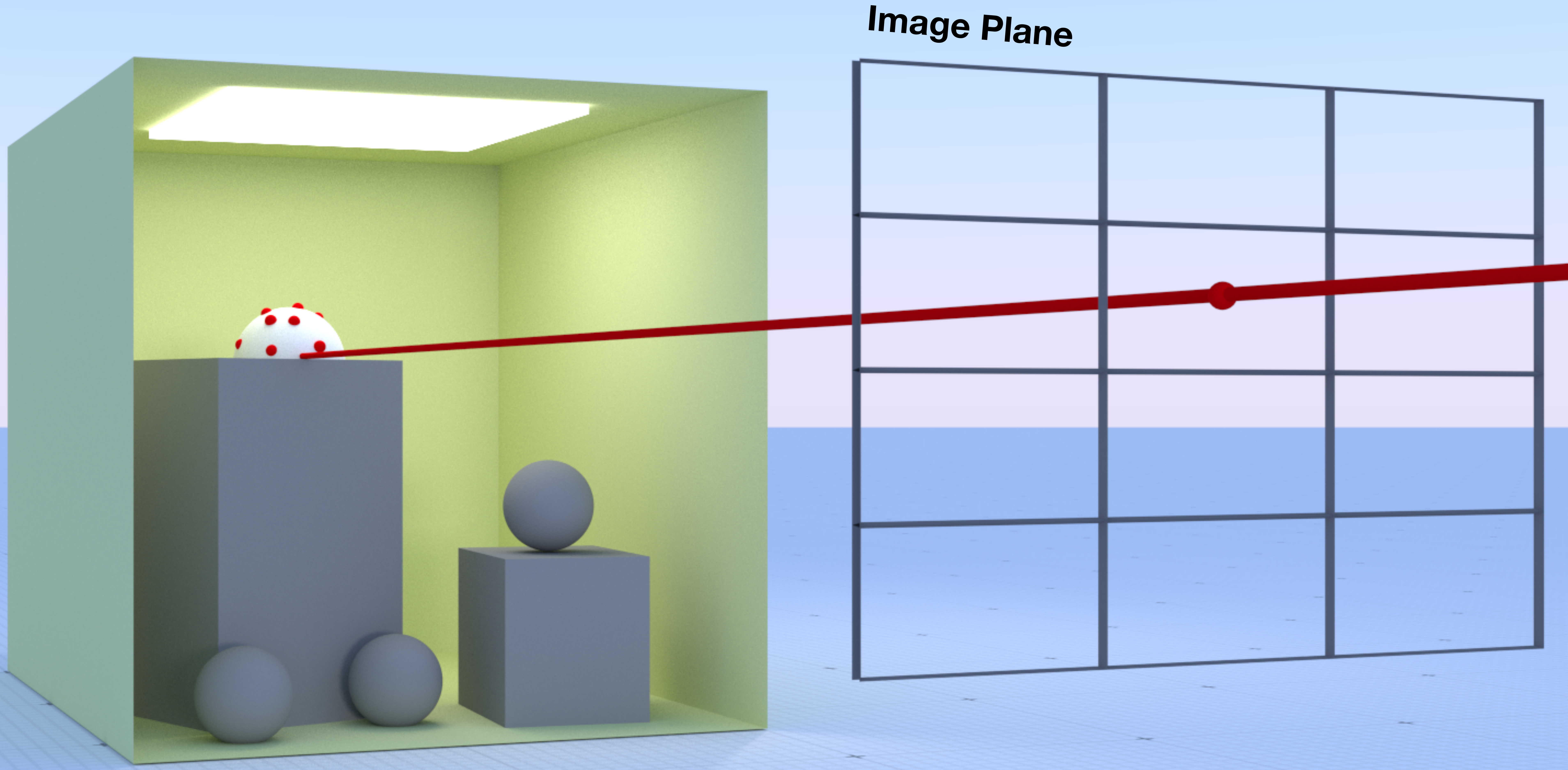
Philipp Slusallek   *Karol Myszkowski*
Gurprit Singh

A la Carte

- σ-algebra and measure
- Random Variables
- Probability distribution functions (PDFs and PMFs)
  - Conditional and Marginal PDFs
- Expected value and Variance of a random variable
- Monte Carlo Integration

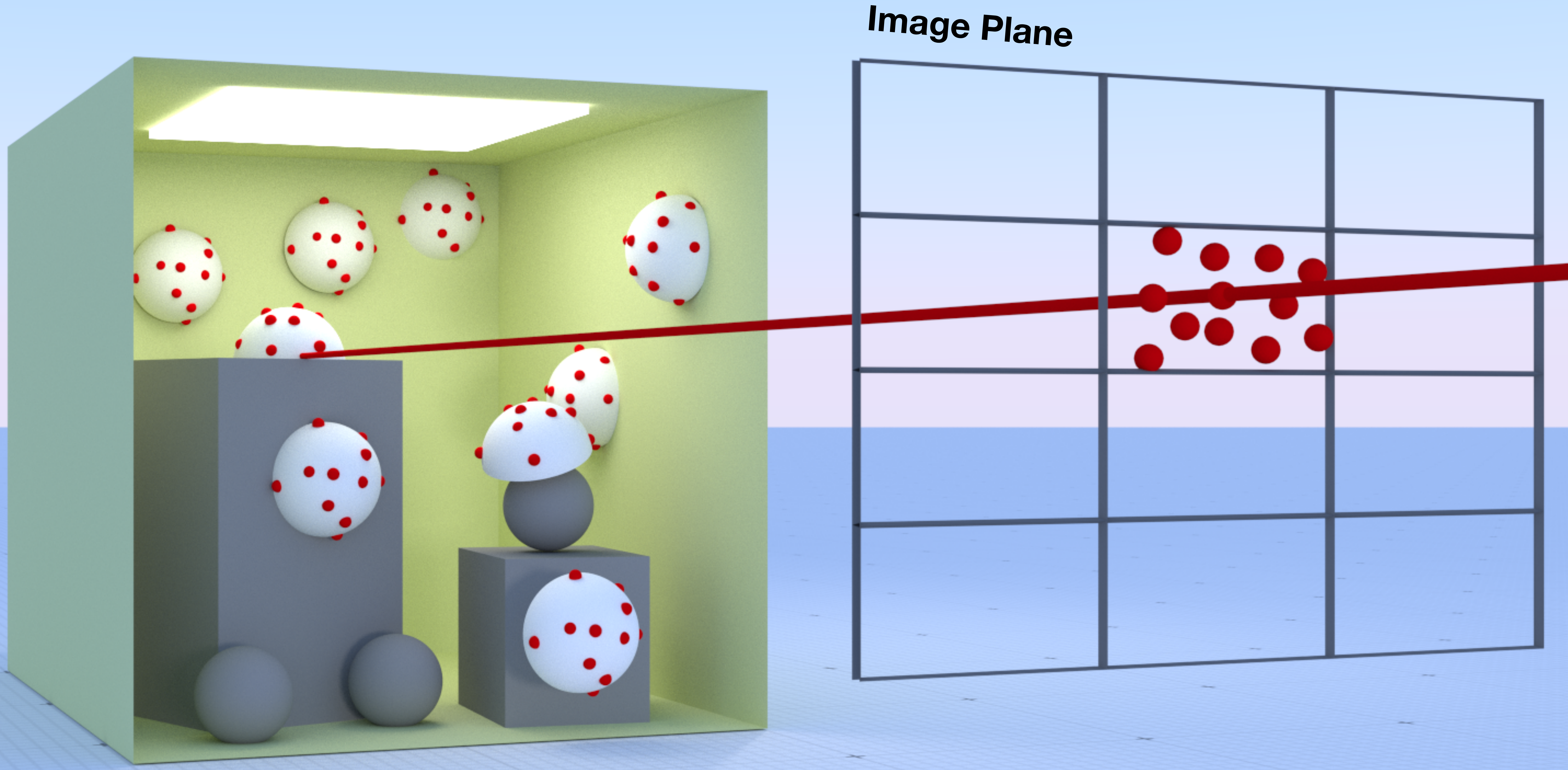
Motivation: Ray Tracing

Ray Tracing



Scene designed by David Coeurjolly

Ray Tracing



Scene designed by David Coeurjolly

Direct Illumination

4 spp

Image rendered using PBRT

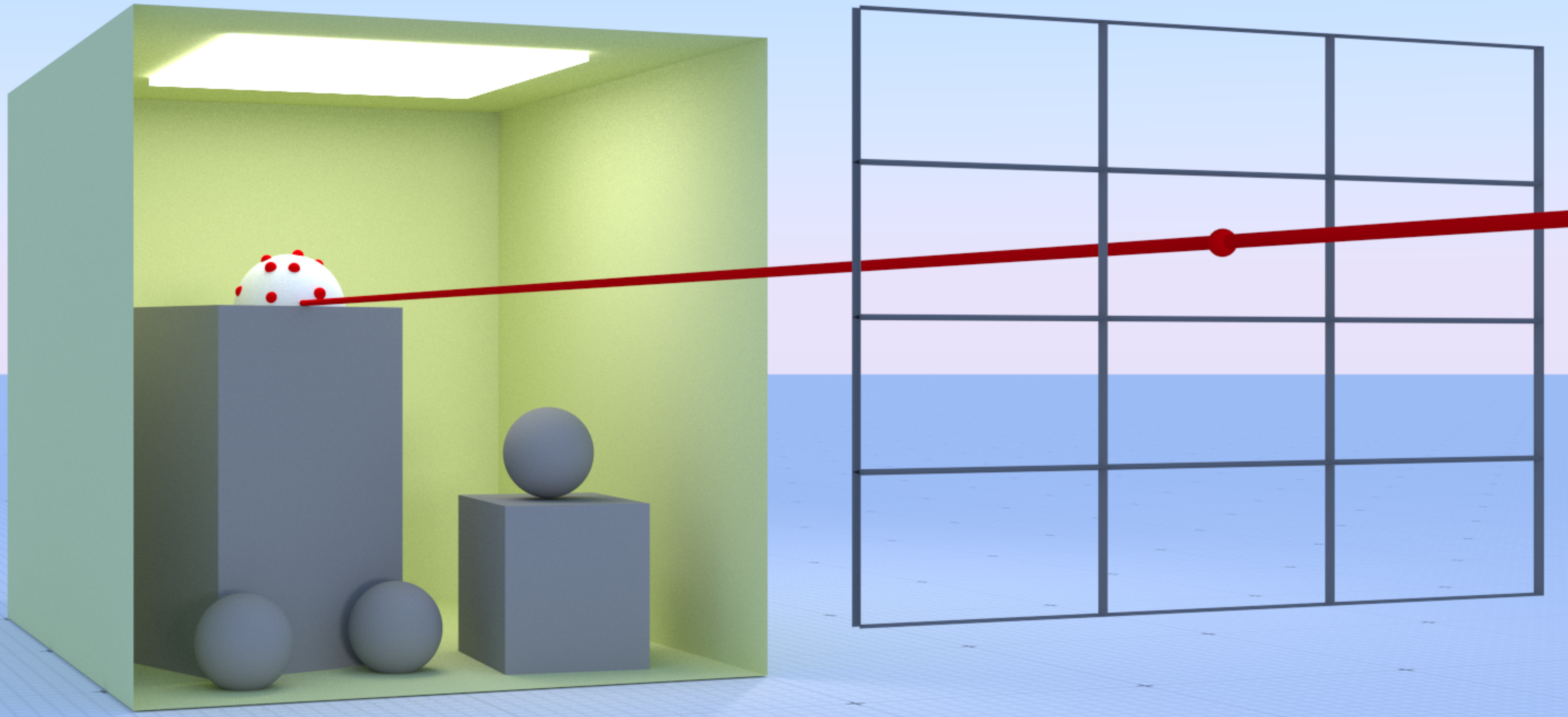
Direct Illumination

256 spp

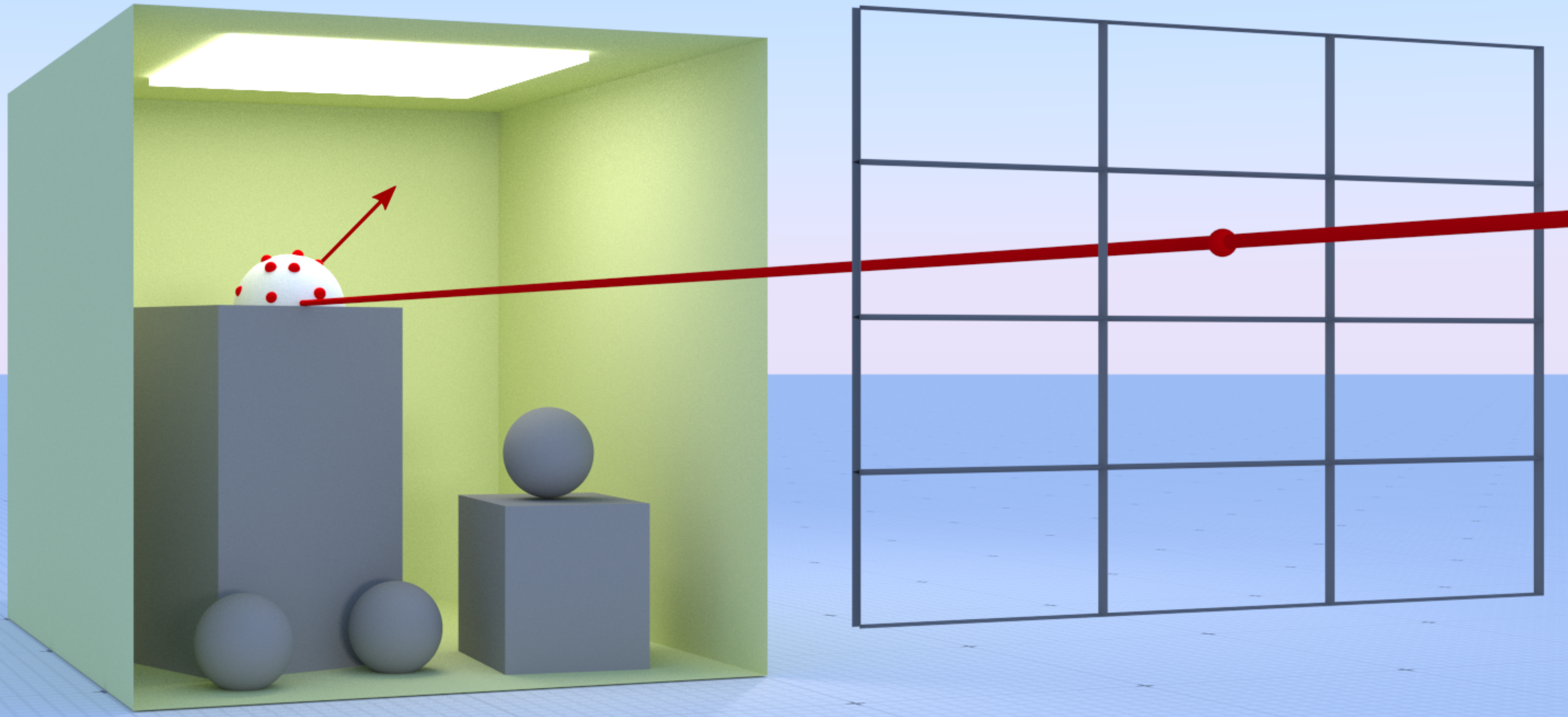
Image rendered using PBRT

Direct and Indirect Illumination

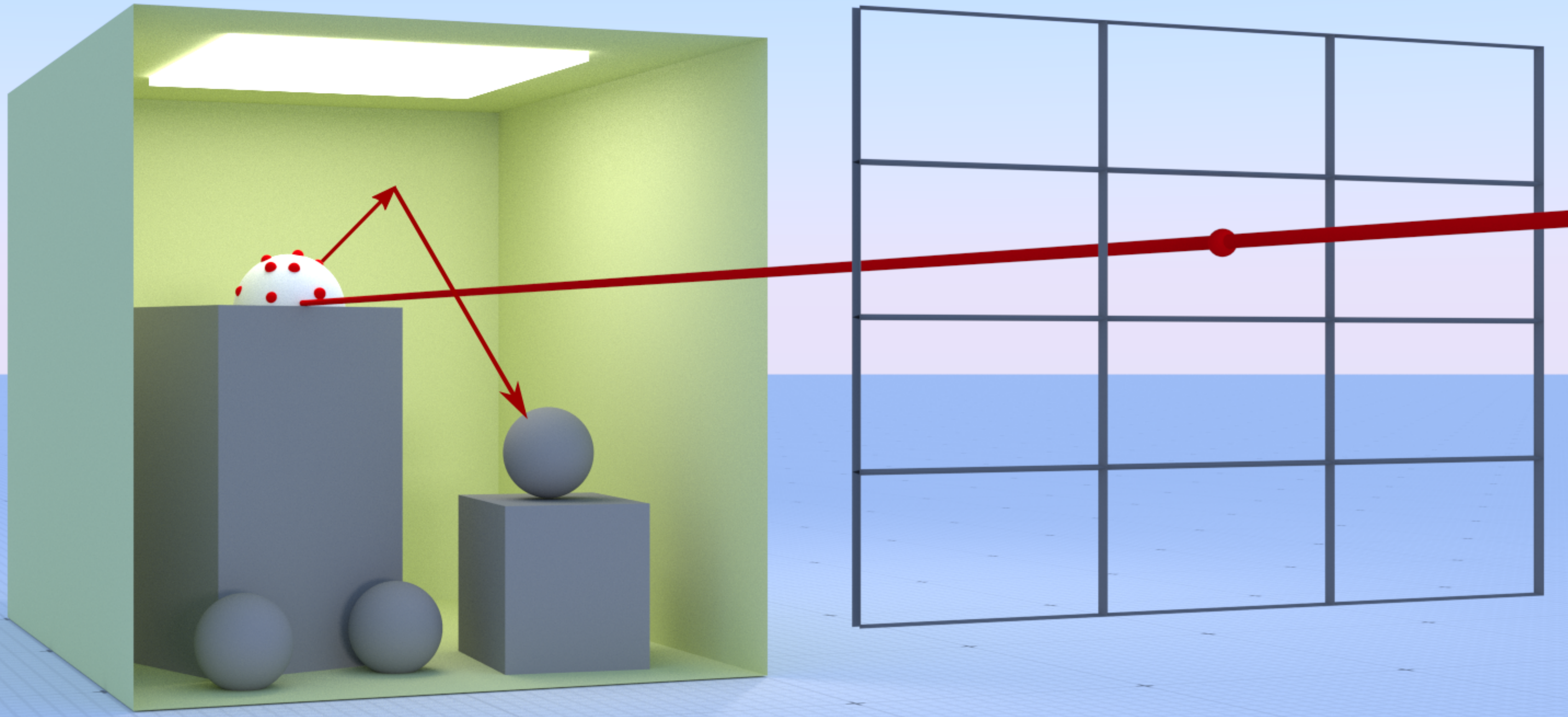
Path Tracing



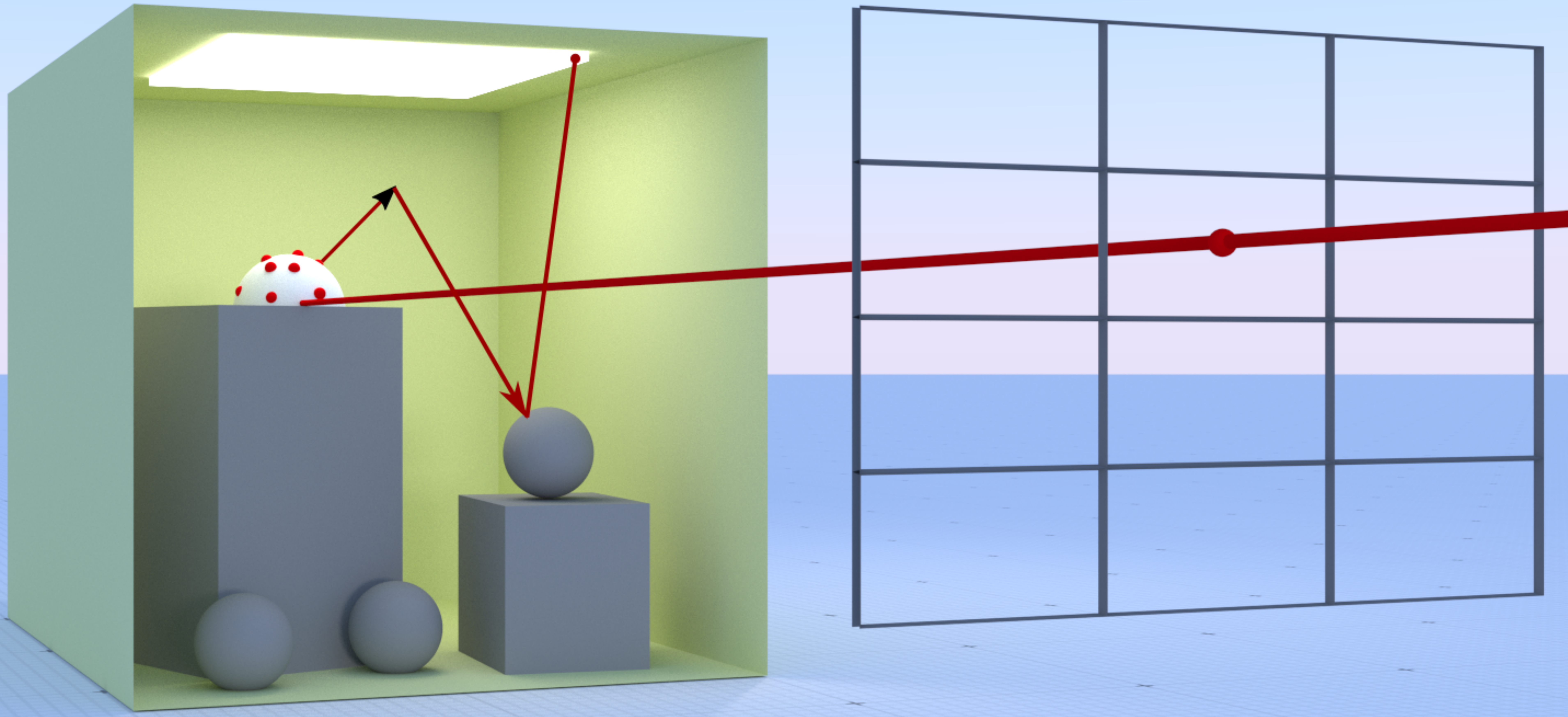
Path Tracing



Path Tracing



Path Tracing



Direct and Indirect Illumination

How can we analyze the noise present in the images ?

Probability Theory
and/or
Number Theory

Probability Theory

- Discrete Probability Space
- Continuous Probability Space

Rolling a fair dice

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Finite outcomes: **discrete** random experiment
- Can ask the outcome is a number: 1 or 6
- Can ask the outcome is a subset, e.g. all prime numbers:

Rolling a fair dice

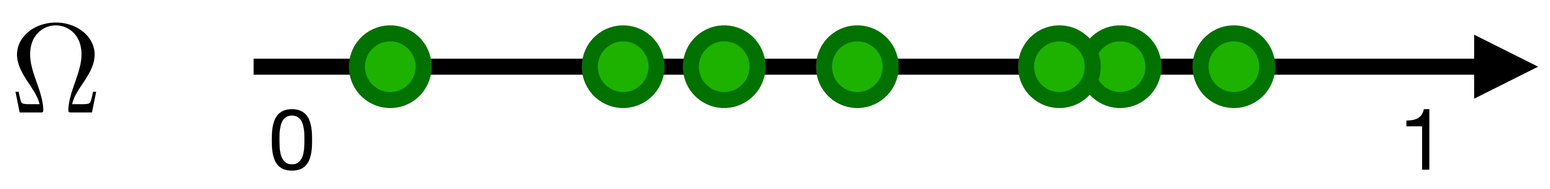
$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- **R1**: Apart from elementary values, the focus lies on subsets of Ω
- **R2**: A probability assigns each element or each subset of a positive real value

The first requirement leads to the concept of σ-algebra

The second to the mathematical construct of a measure

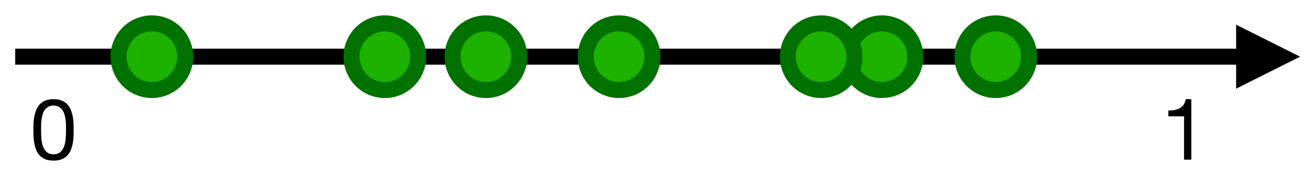
Random number in [0,1]



- Uncountably infinite outcomes: **continuous** random experiment
- Does not make sense to ask for one number as output, e.g. 0.245
- We need to ask for the probability of a region, e.g. [0.2,0.4] or [0.36,0.89]

Random number in [0,1]

Ω



- **R1**: As in discrete case, focus lies on subsets of Ω, also called events
- **R2**: A probability assigns each subset of a positive real value.

The first requirement leads to the concept of Borel σ-algebra

The second to the mathematical construct of a Lebesgue measure

σ-Algebra

- Mathematical construct used in probability and measure theory

   1. Take on the role of system of events in probability theory

- Simply spoken: Collection of subsets of a given set

   A. A non-empty collection of subsets c hat is **closed** under the set theoretical operations of: countable unions, countable intersections, and complement

UNIVERSITÄT DES SAARLANDES

σ-Algebra

- For discrete set Ω :

    1. The sigma-algebra corresponds to the power set of omega (set of all subsets)

σ-Algebra

- For discrete set Ω :

    1. The sigma-algebra corresponds to the power set of omega (set of all subsets)

$$\Omega = \{0, 1\}$$

$$\Sigma = \{\{\phi\}, \{0\}, \{1\}, \{0, 1\}\}$$

σ-Algebra

- For discrete set Ω :

    1. The sigma-algebra corresponds to the power set of omega (set of all subsets)

$$\Omega = \{0, 1\}$$

$$\Sigma = \{\{\phi\}, \{0\}, \{1\}, \{0, 1\}\}$$

$$\Omega = \{a, b, c, d\}$$

$$\Sigma = \{\{\phi\}, \{a, b\}, \{c, d\}, \{a, b, c, d\}\}$$

σ-Algebra

- For continuous set Ω:

A. The associated sigma algebras are the Borel sets ove   i.e., the collection of all open sets over omega that can be generated via countable unions, countable intersections, and complement of open sets

σ-Algebra

- For continuous set Ω:

A. The associated sigma algebras are the Borel sets over Ω, i.e., the collection of all open sets over omega that can be generated via countable unions, countable intersections, and complement of open sets

$I = [p, q), p, q \in \mathbb{R}$ Fixed half-interval

σ-Algebra

- For continuous set Ω:

A. The associated sigma algebras are the Borel sets over Ω, i.e., the collection of all open sets over omega that can be generated via countable unions, countable intersections, and complement of open sets

$I = [p, q), p, q \in \mathbb{R}$ Fixed half-interval

$\mathbb{T} = [\alpha, \beta) \subseteq [p, q)$ Collection of all half-intervals

σ-Algebra

- For continuous set Ω:

A. The associated sigma algebras are the Borel sets over Ω, i.e., the collection of all open sets over omega that can be generated via countable unions, countable intersections, and complement of open sets

$I = [p, q), p, q \in \mathbb{R}$ Fixed half-interval

$\mathbb{T} = [\alpha, \beta) \subseteq [p, q)$ Collection of all half-intervals

Here, $\mathbb{T}$ is not a σ-algebra because, generally speaking, neither the union nor the difference of two half-intervals is a half-interval.

σ-Algebra

It is the mathematical construct that allows defining a measure

Measure

- In probability theory, it plays the role of a probability distribution
- A real-valued set function defined on a sigma-algebra that assigns each subset of a sigma-algebra a non-negative real number.
- A sigma-additive set function: i.e., the measure of the union of disjoint sets is equal to the sum of the measures of the individual sets

Lebesgue Measure

- Standard way of assigning measure to subsets of n-dimensional Euclidean space.
- For n = 1,2 or 3, it coincides with the standard measure of length, area or volume, respectively.

Random Variable

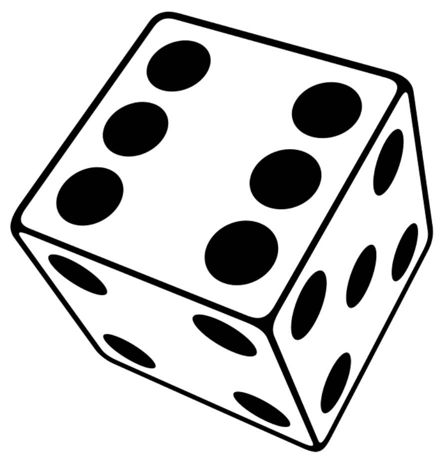
- Central concept in probability theory
- Enables to construct a simpler probability space from a rather complex one
- Correspond to a measurable function defined on a gebra that assigns each element to a real number

Random Variable

- A random variable X is a value chosen by some random process
- Random variables are always drawn from a domain: discrete (e.g., a fixed set of probabilities) or continuous (e.g., real numbers)
- Applying a func random var ults in a new random variable

Discrete Probability Space

Discrete Random Variable



- Random variable (RV):

$$X : \Omega \to E \qquad\qquad \Omega = \{x_1, x_2, \ldots, x_n\}$$

- Probabilities:

$$\{p_1, p_2, \ldots, p_n\}$$

$$\sum_{i=1}^{N} p_i = 1$$

Discrete Random Variable

- Example: Rolling a Die

$$x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5, x_6 = 6$$

- Probability of each event:

$p_i = 1/6 \quad \text{for } \; i = 1, \ldots, 6$ $\qquad P(X = i) = \frac{1}{6}$

Discrete Random Variable

$$P(2 \leq X \leq 4) = \sum_{i=2}^{4} P(X = i)$$

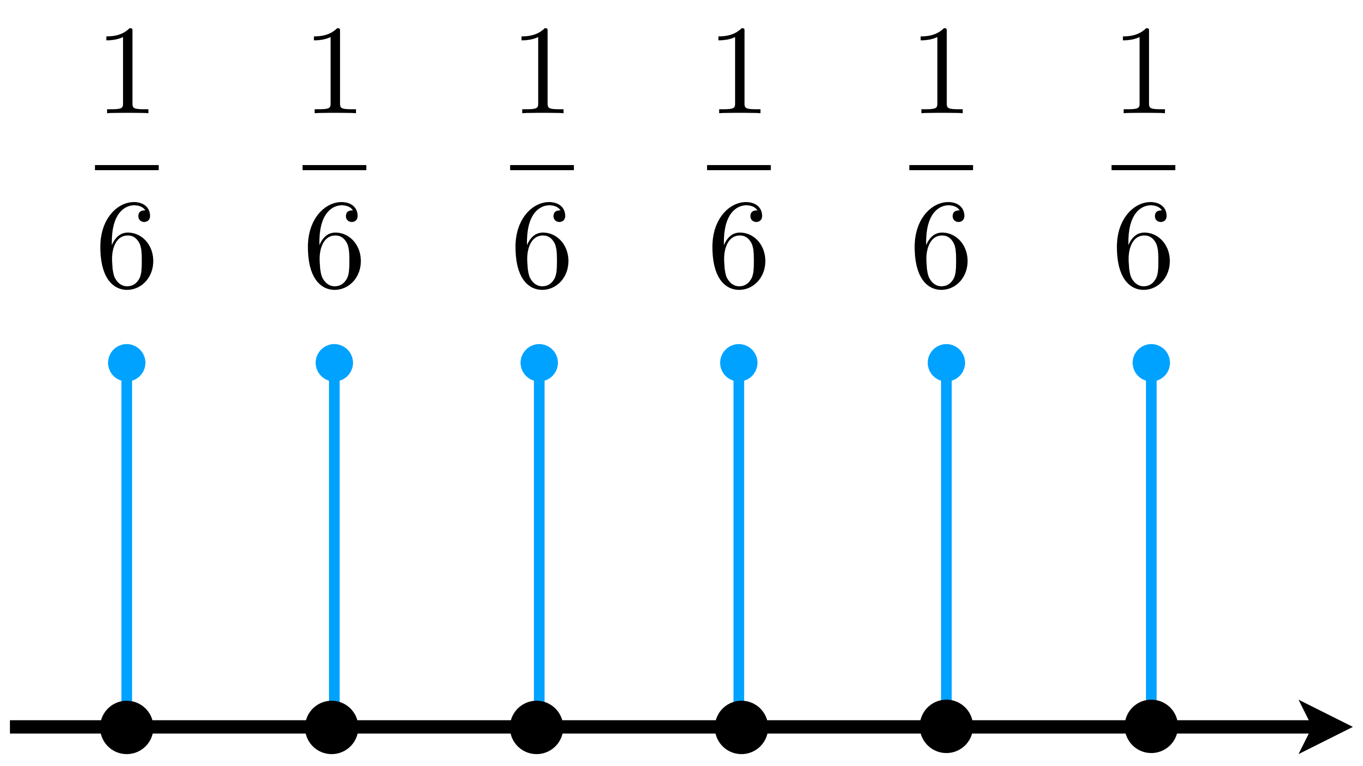
$$= \sum_{i=2}^{4} \frac{1}{6} = \frac{1}{2}$$

Probability mass function

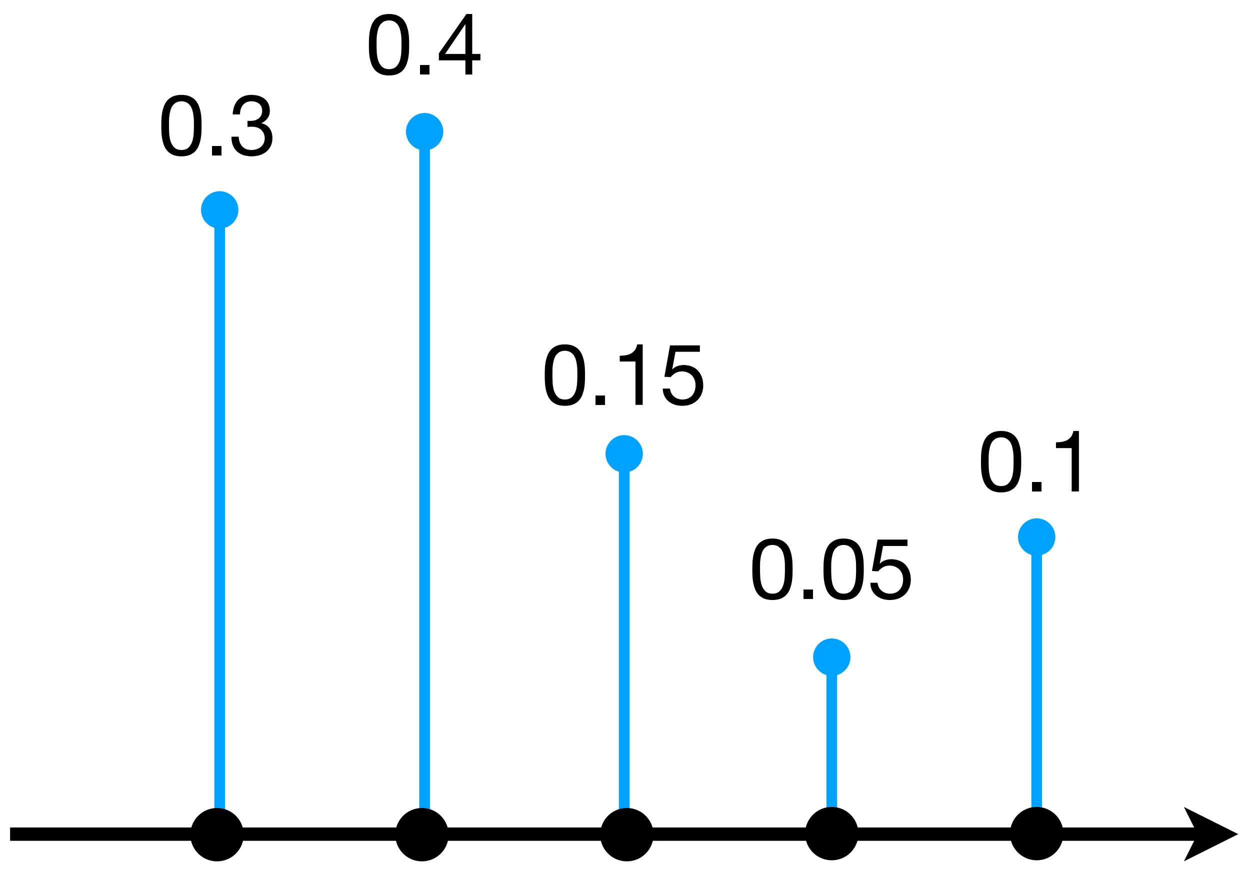
- PMF is a function that gives the probability that a discrete RV is exactly equal to some value.
- PMF is different from PDF (probability density function) which is for continuous RVs.

Probability mass function

Constant PMF



Non-uniform PMF

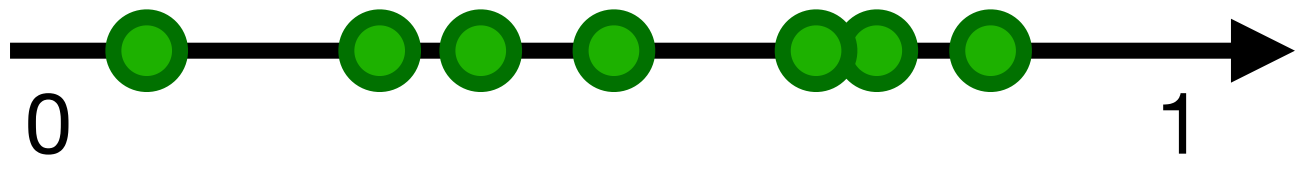


Continuous Probability Space

Continuous Random Variable

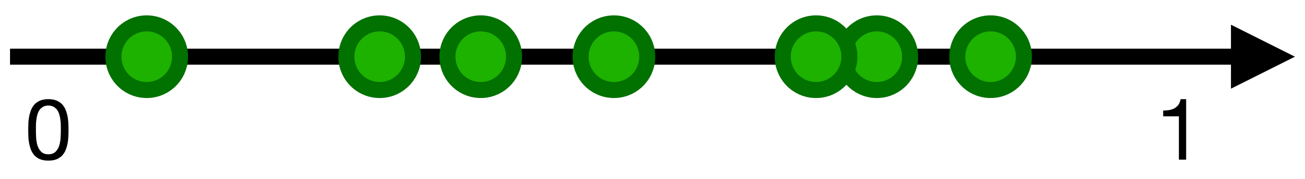
- In rendering, discrete random variables are less common than continuous random variables
- Continuous random variables take on values that ranges of continuous domains (e.g. real numbers or directions on the unit sphere)
- A particularly important random variable is the *canonical uniform random variable, which we write as* ξ

Continuous Random Variable



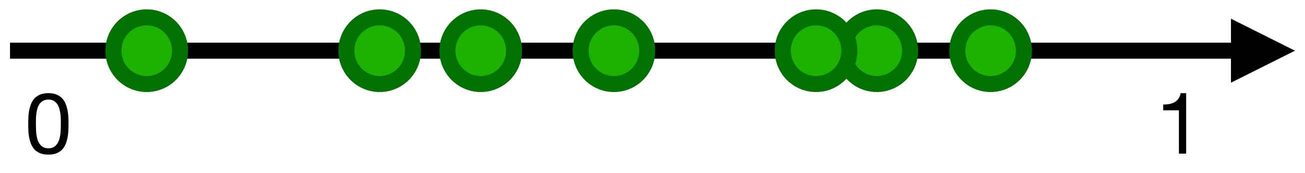
$$\xi \in [0, 1)$$

Continuous Random Variable



- We can take a continuous, uniformly distributed random variable $\xi \in [0, 1)$ and map to a discrete random variable, choosing X_i if:

Continuous Random Variable



- We can take a continuous, uniformly distributed random variable $\xi \in [0, 1)$ and map to a discrete random variable, choosing X_i if:

$$\sum_{j=1}^{i-1} p_j < \xi \leq \sum_{j=1}^{i} p_j$$

$$X_i = \{1, 2, 3, 4, 5, 6\}$$

Visual Break

Image rendered using PBRT

Visual Break

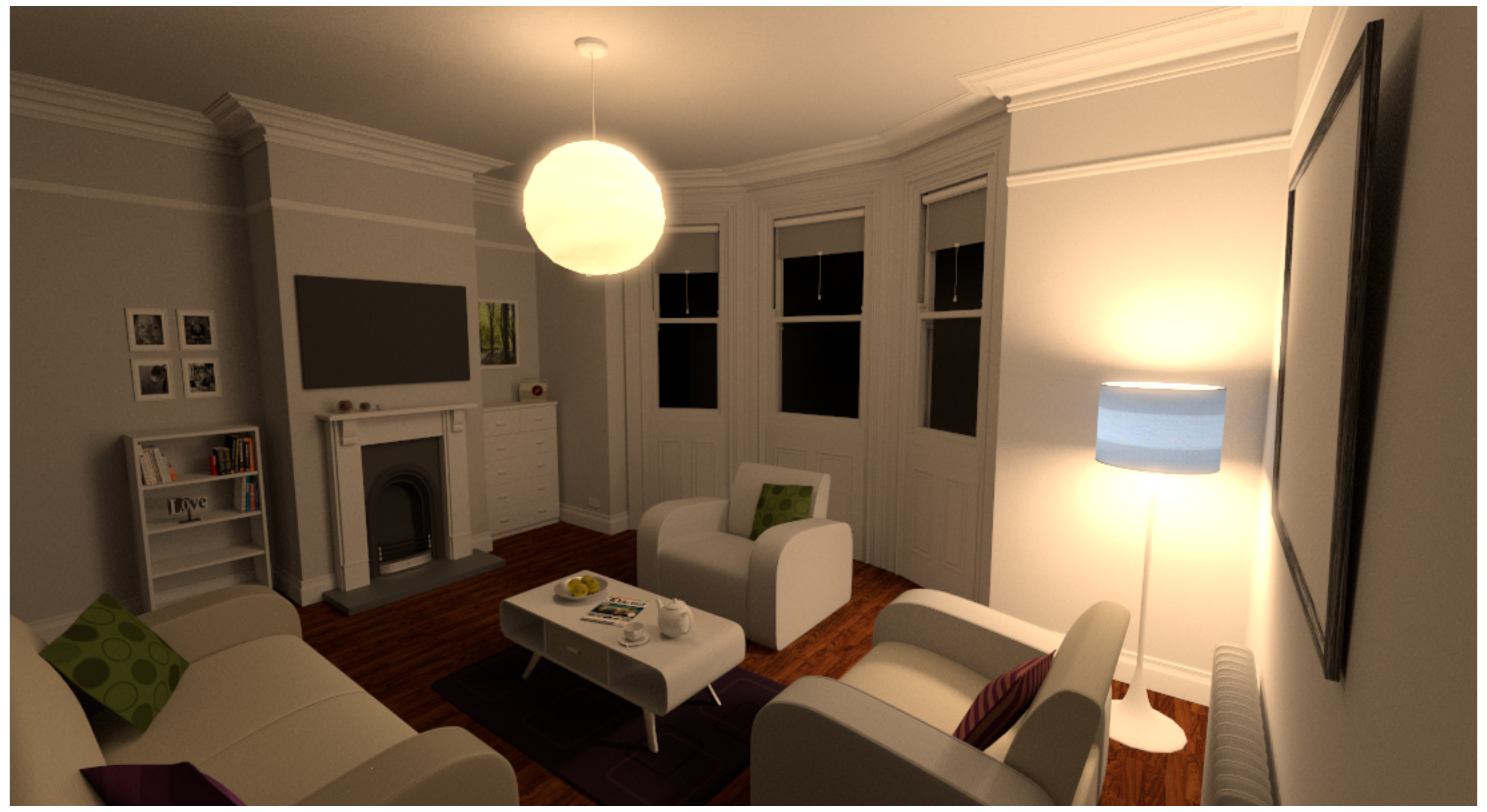
Image rendered using PBRT

Continuous Random Variable

- For lighting application, we might want to define probability of sampling illumination from each light source in the scene based on its power Φ_i

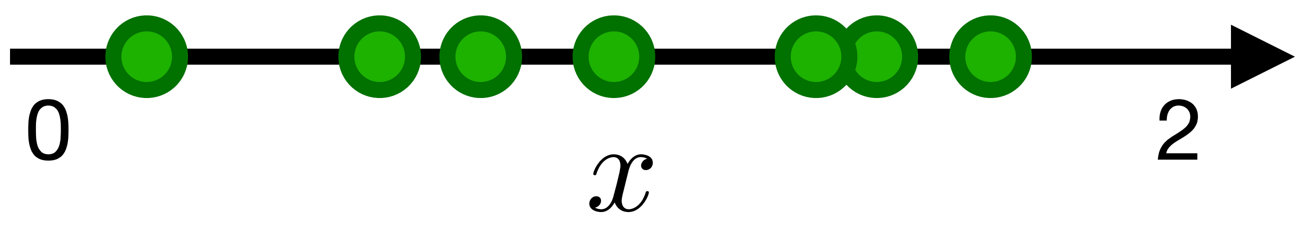
$$p_i = \frac{\Phi_i}{\sum_j \Phi_j}$$

Here, the probability is relative to the total power



Probability Density Functions

Probability density function



- Consider a continuous RV that ranges over real numbers: $[0, 2)$, where the probability of taking on any particular value x is **proportional** to the value $2 - x$

- It is twice as likely for this random variable to take on a value around 0 as it is to take around 1, and so forth.

Probability density function

- The probability density function (PDF) formalizes this idea: it describes the relative probability of a RV taking on a particular value.
- Unlike PMF, the values of the PDFs are not the probabilities as such: a PDF must be integrated over an interval to yield a probability

Probability density function

For uniform random variables:

$$p(x) = \begin{cases} 1 & x \in [0, 1) \\ 0 & \text{otherwise} \end{cases}$$

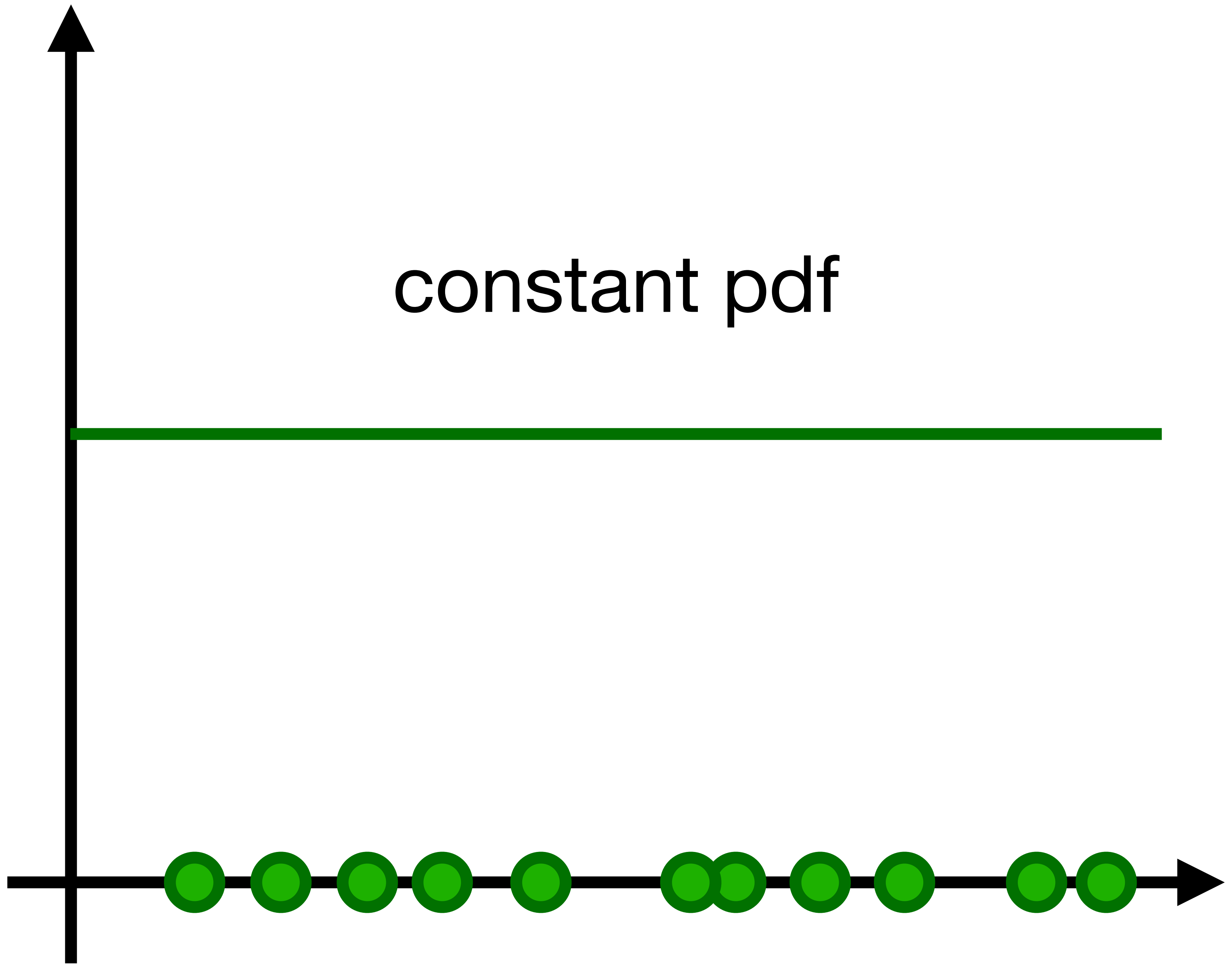
For non-uniform random variables:

$p(x)$ could be any function

Probability density function

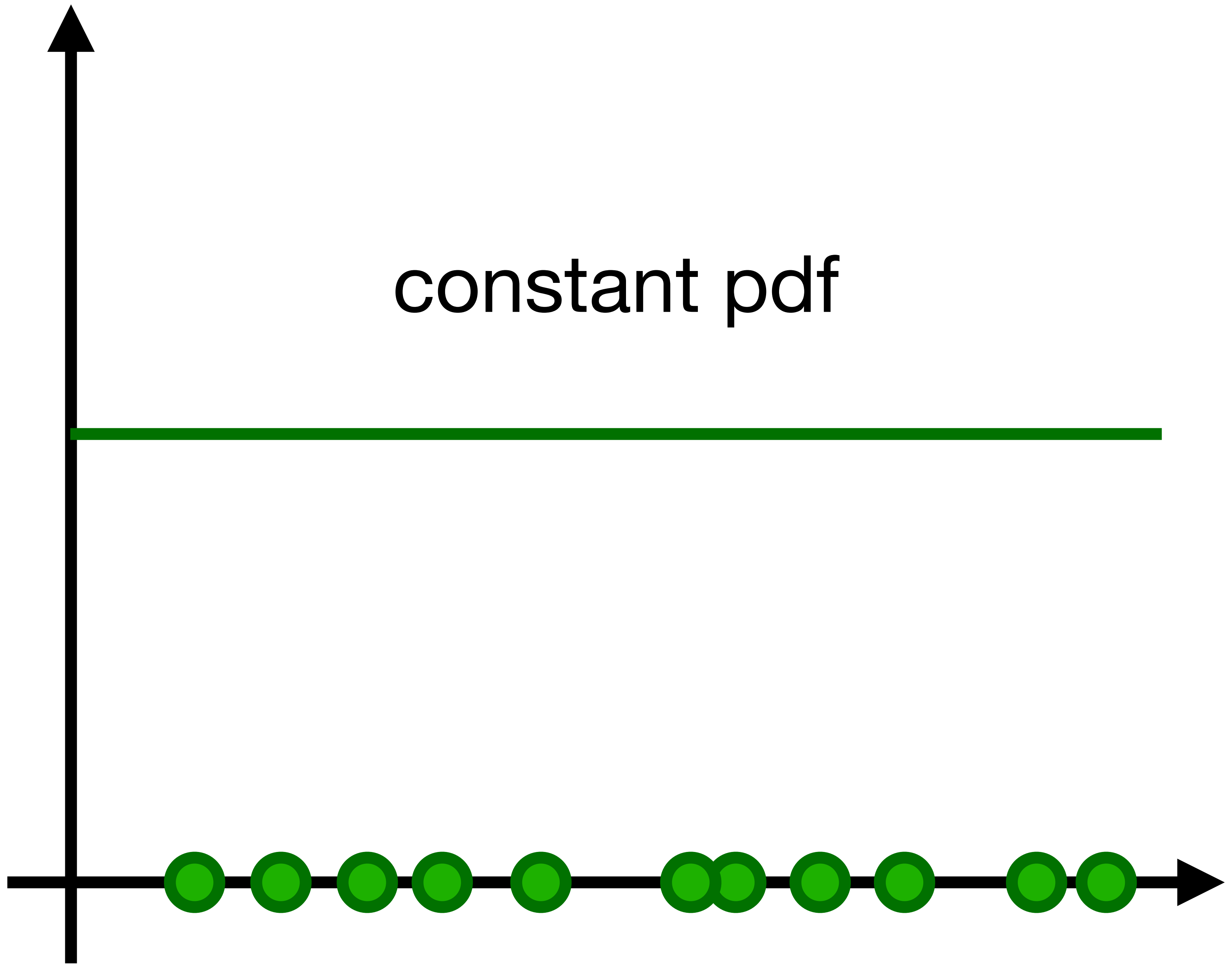
Uniform distribution

Non-uniform distribution

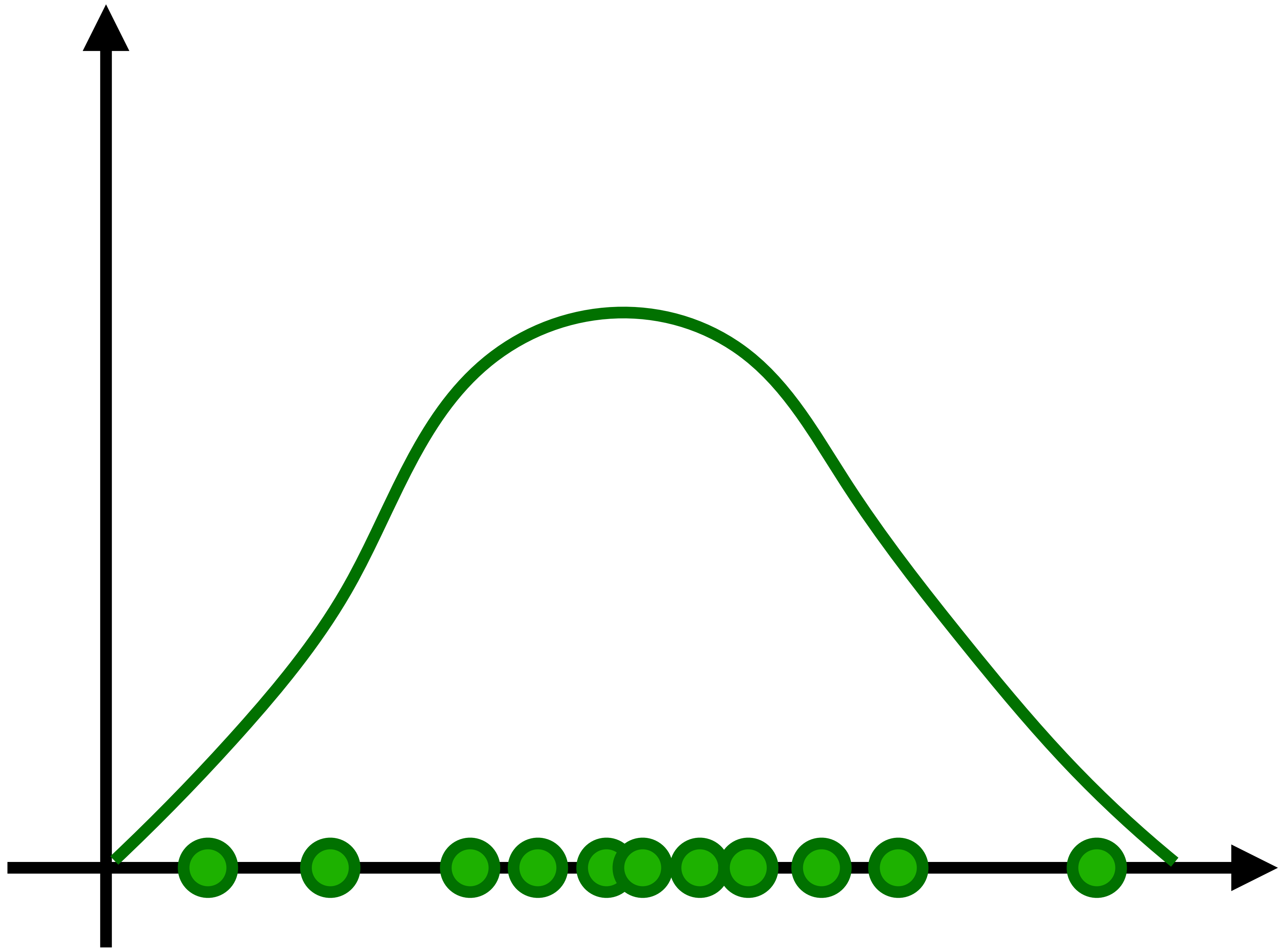


Probability density function

Uniform distribution



Non-uniform distribution

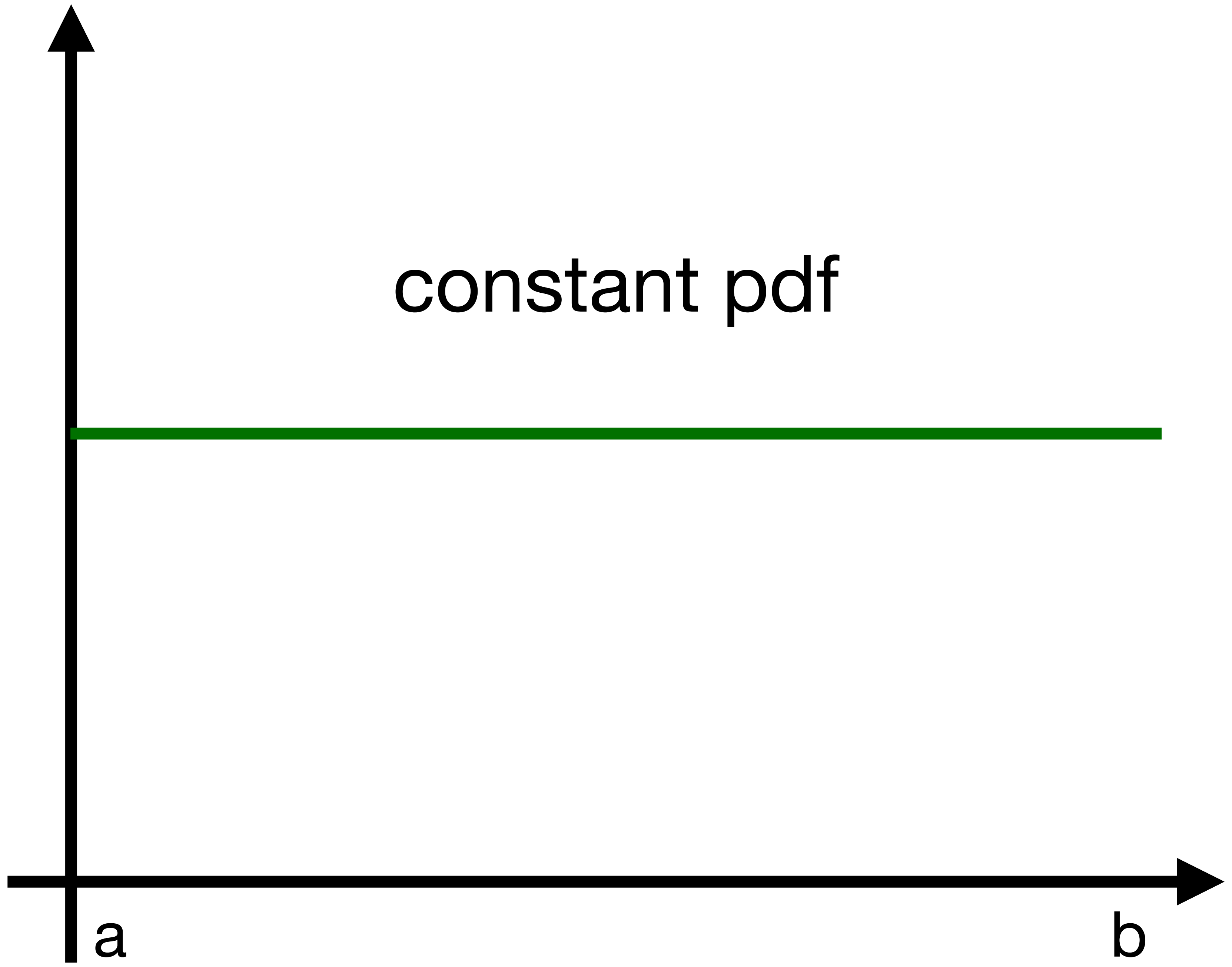


Probability density function

Some properties of PDFs: $$p(x) > 0$$

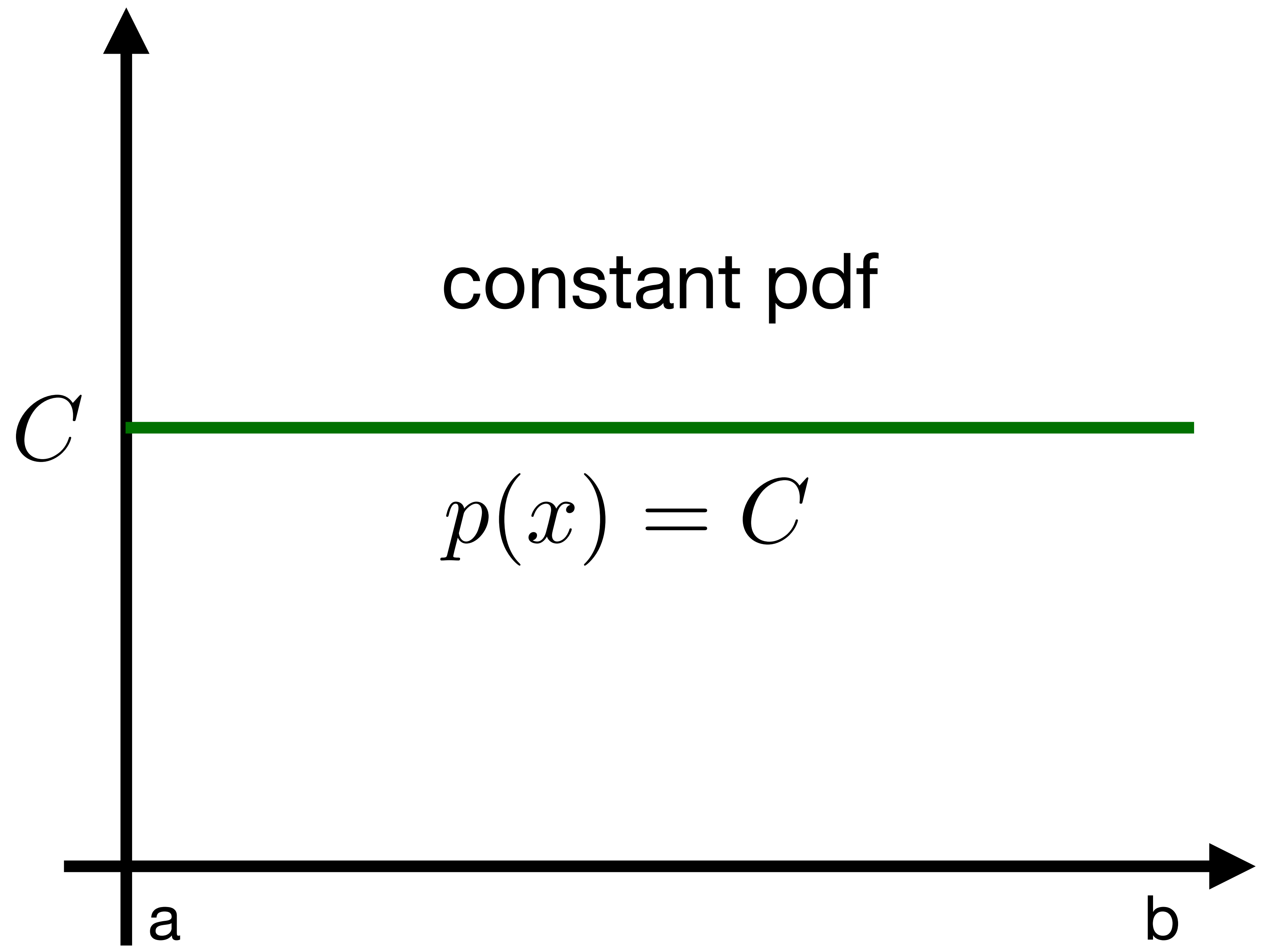
$$\int_{-\infty}^{\infty} p(x)dx = 1$$

Probability density function



$$\int_a^b p(x)dx = 1 \quad x \in [a, b)$$

Probability density function



$$\int_a^b p(x)dx = 1 \qquad x \in [a, b)$$

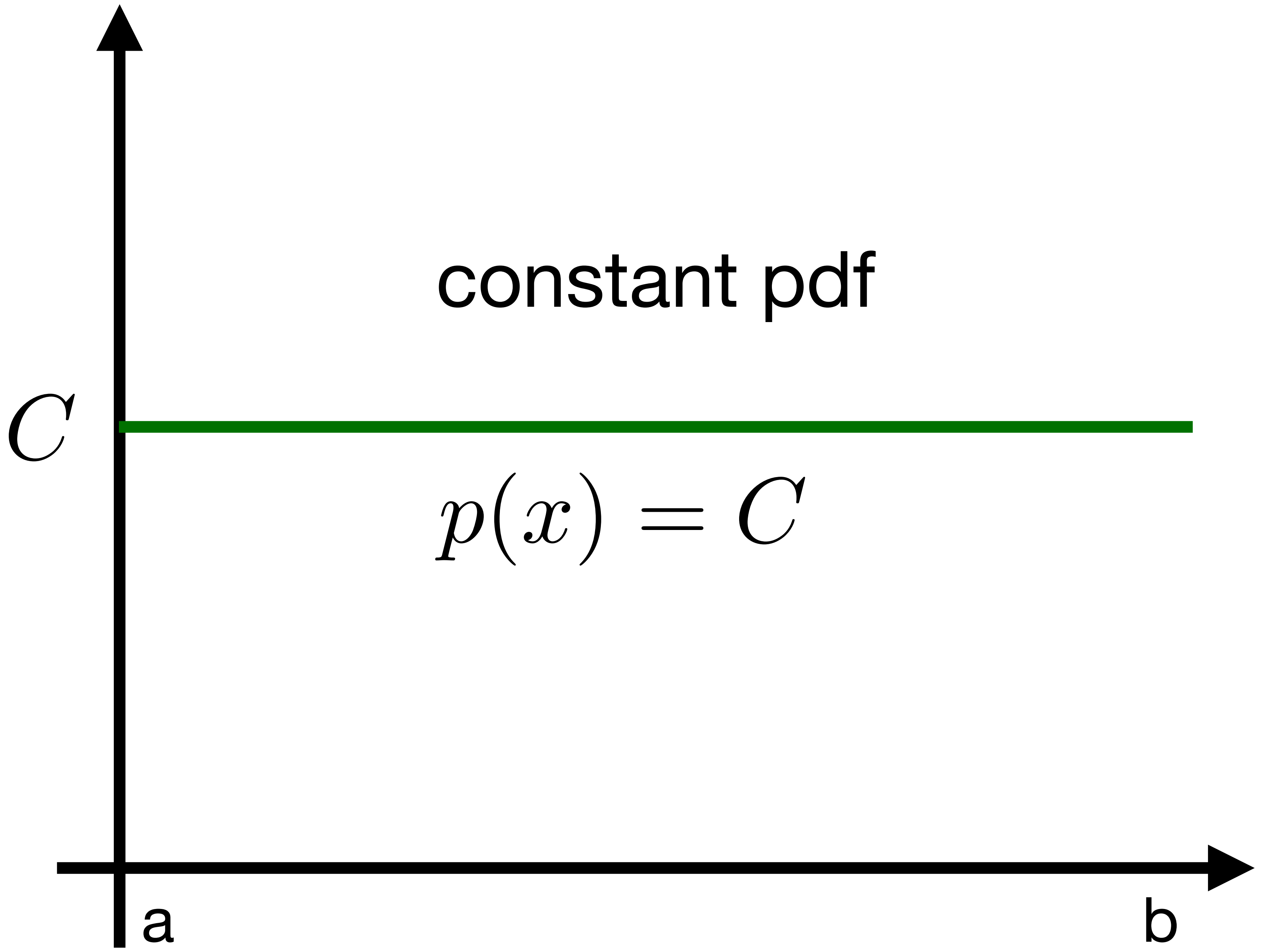
$$\int_a^b C\, dx = 1$$

$$C \int_a^b dx = 1$$

$$C(b-a) = 1$$

$$C = \frac{1}{b-a}$$

Probability density function



$$\int_a^b p(x)dx = 1 \qquad x \in [a, b)$$

$$\int_a^b C \; dx = 1$$

$$C \int_a^b dx = 1$$

$$C(b-a) = 1$$

$$C = \frac{1}{b-a}$$

$$\boxed{p(x) = \frac{1}{b-a}}$$

Cumulative distribution function

- The PDF $p(x)$ is the derivative of the random variable's CDF:

Cumulative distribution function

- The PDF $p(x)$ is the derivative of the random variable's CDF:

$$p(x) = \frac{dP(x)}{dx}$$

$P(x)$: cumulative distribution function (CDF) ,
also called cumulative density function

Cumulative distribution function

- The PDF $p(x)$ is the derivative of the random variable's CDF:

$$p(x) = \frac{dP(x)}{dx}$$

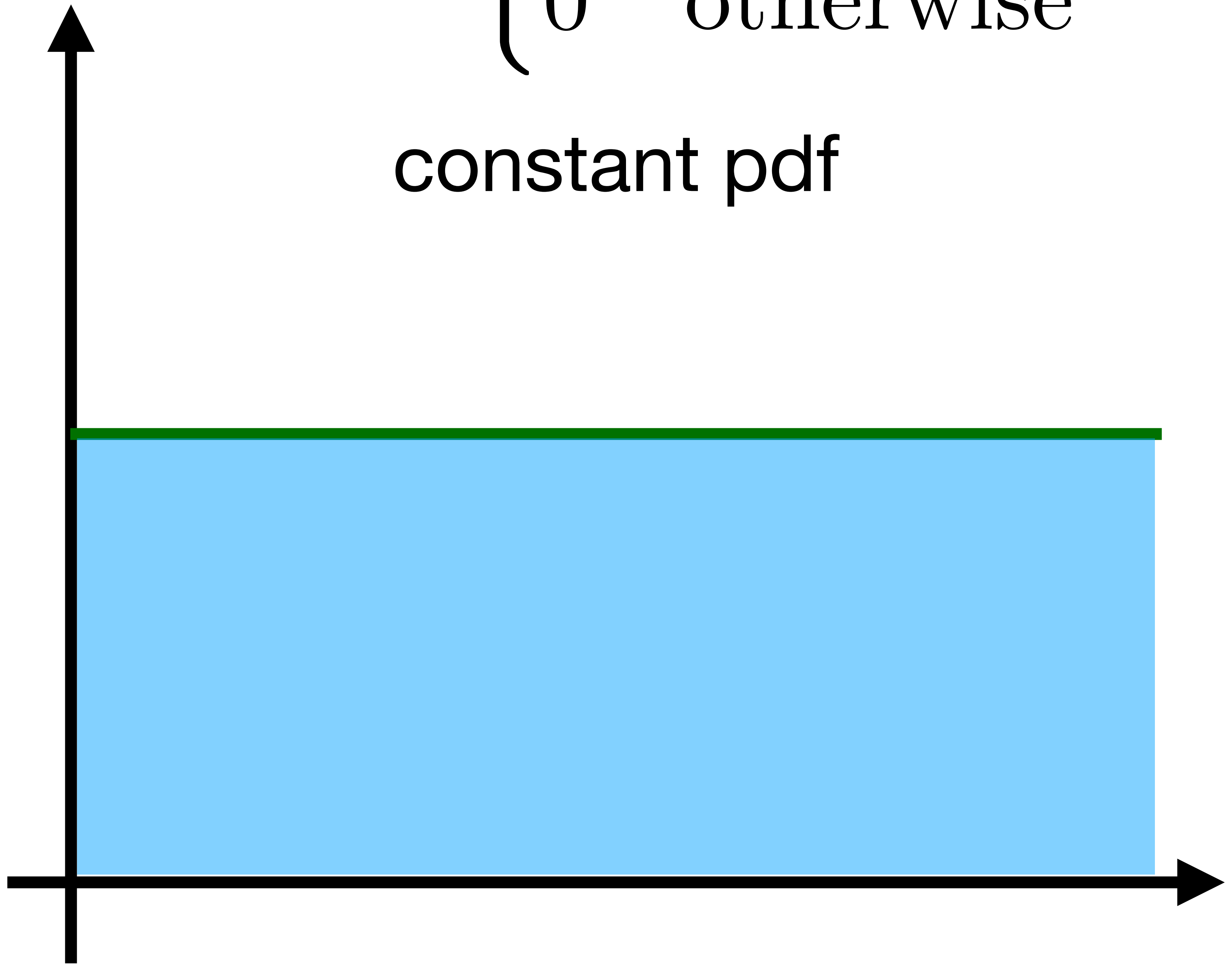
$$P(x) = \int_{-\infty}^{x} p(x)dx$$

$P(x)$: cumulative distribution function (CDF) ,
also called cumulative density function

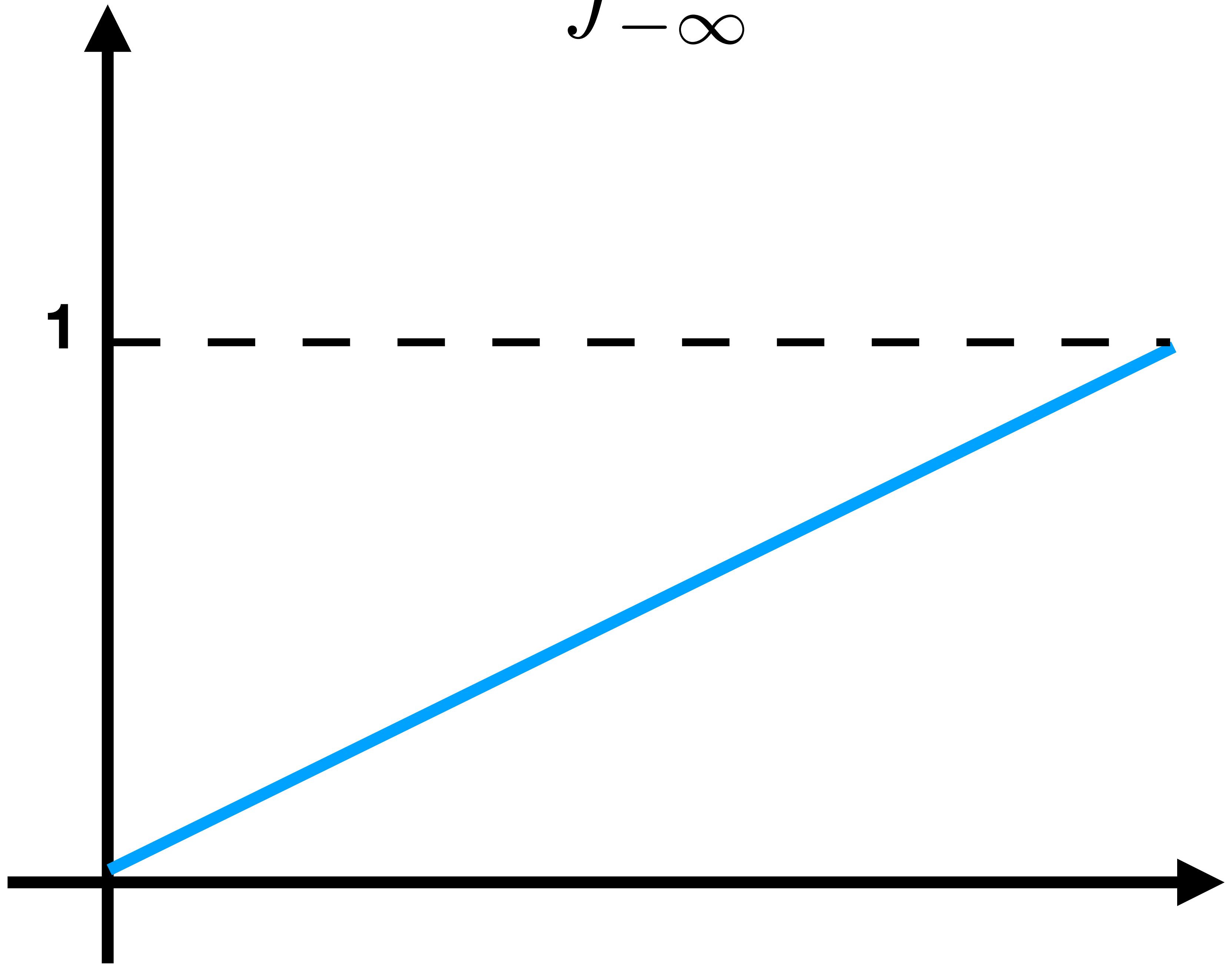
Cumulative distribution function

$$p(x) = \begin{cases} 1 & x \in [0,1) \\ 0 & \text{otherwise} \end{cases}$$

constant pdf

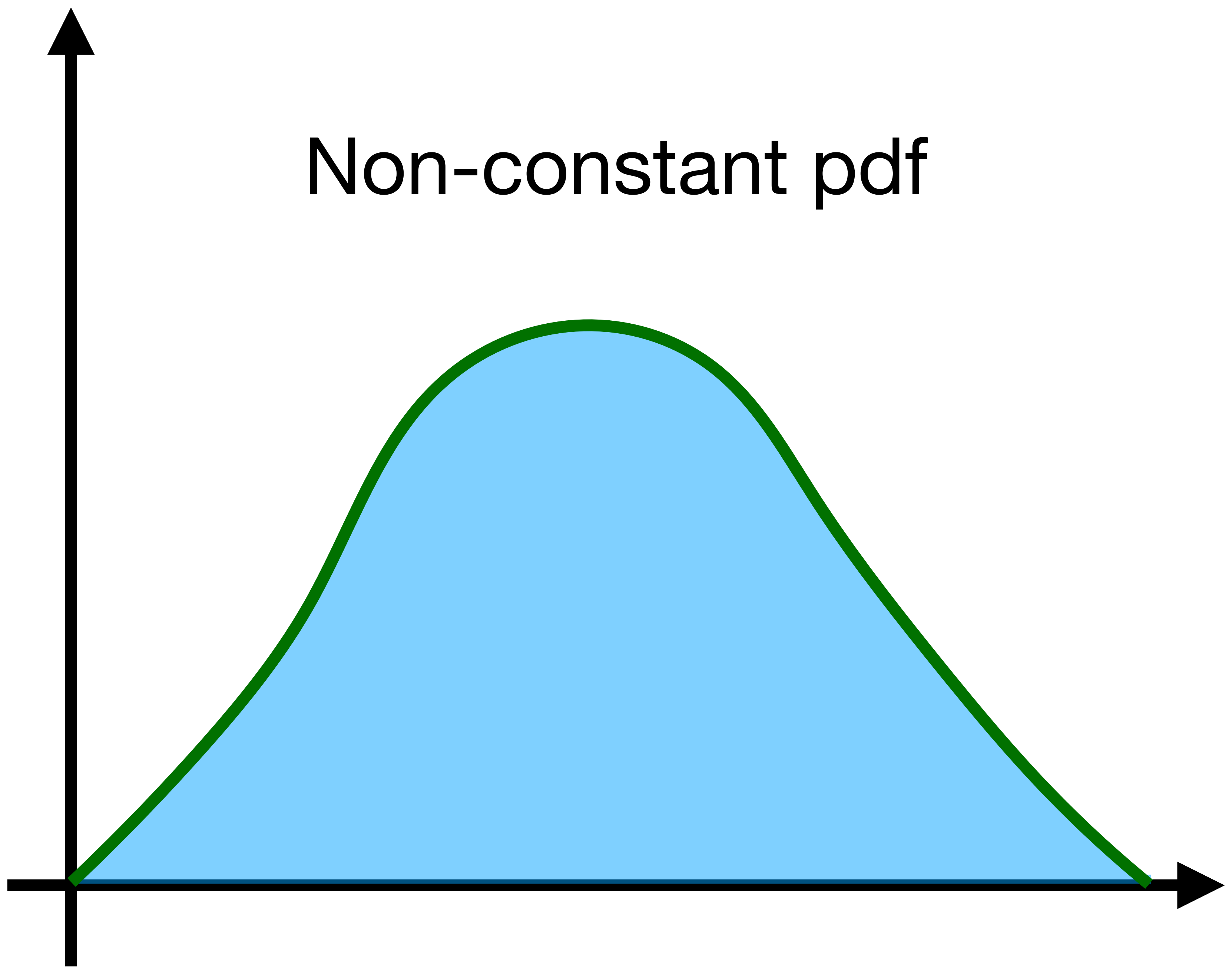


$$P(x) = \int_{-\infty}^{x} p(x)dx$$

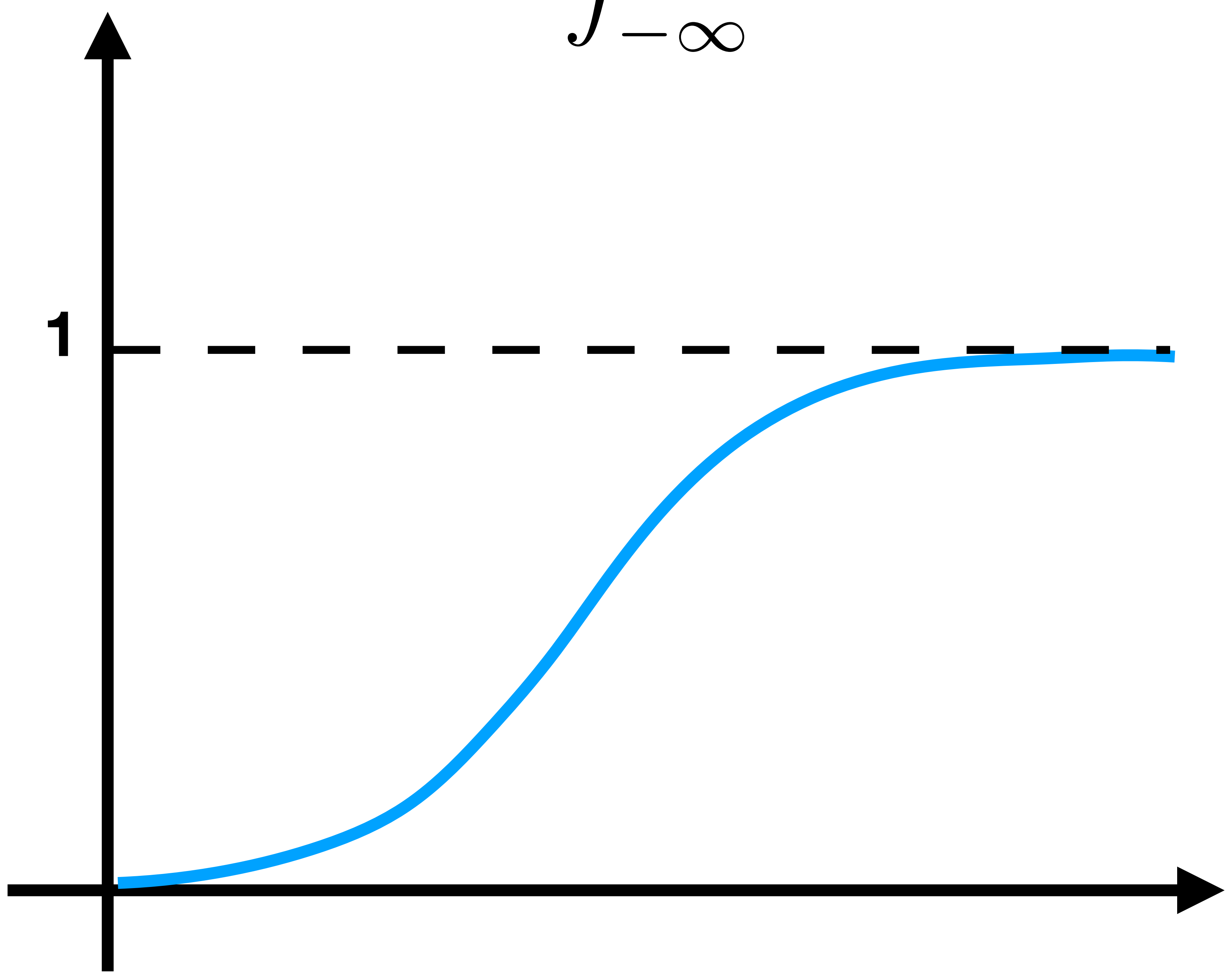


Cumulative distribution function

$p(x)$



$$P(x) = \int_{-\infty}^{x} p(x)dx$$



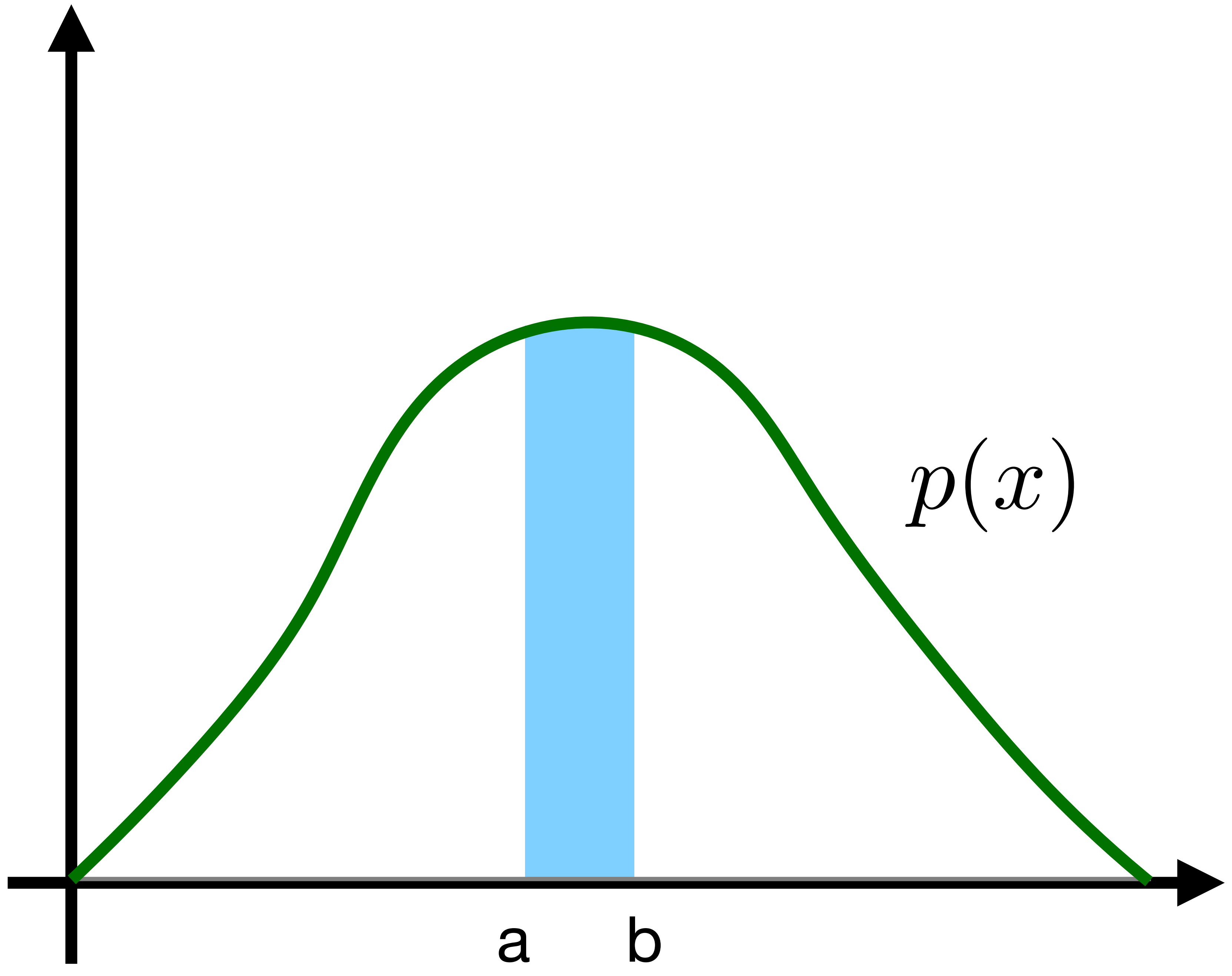
Visual Break

Image rendered using PBRT

Probability: Integral of PDF

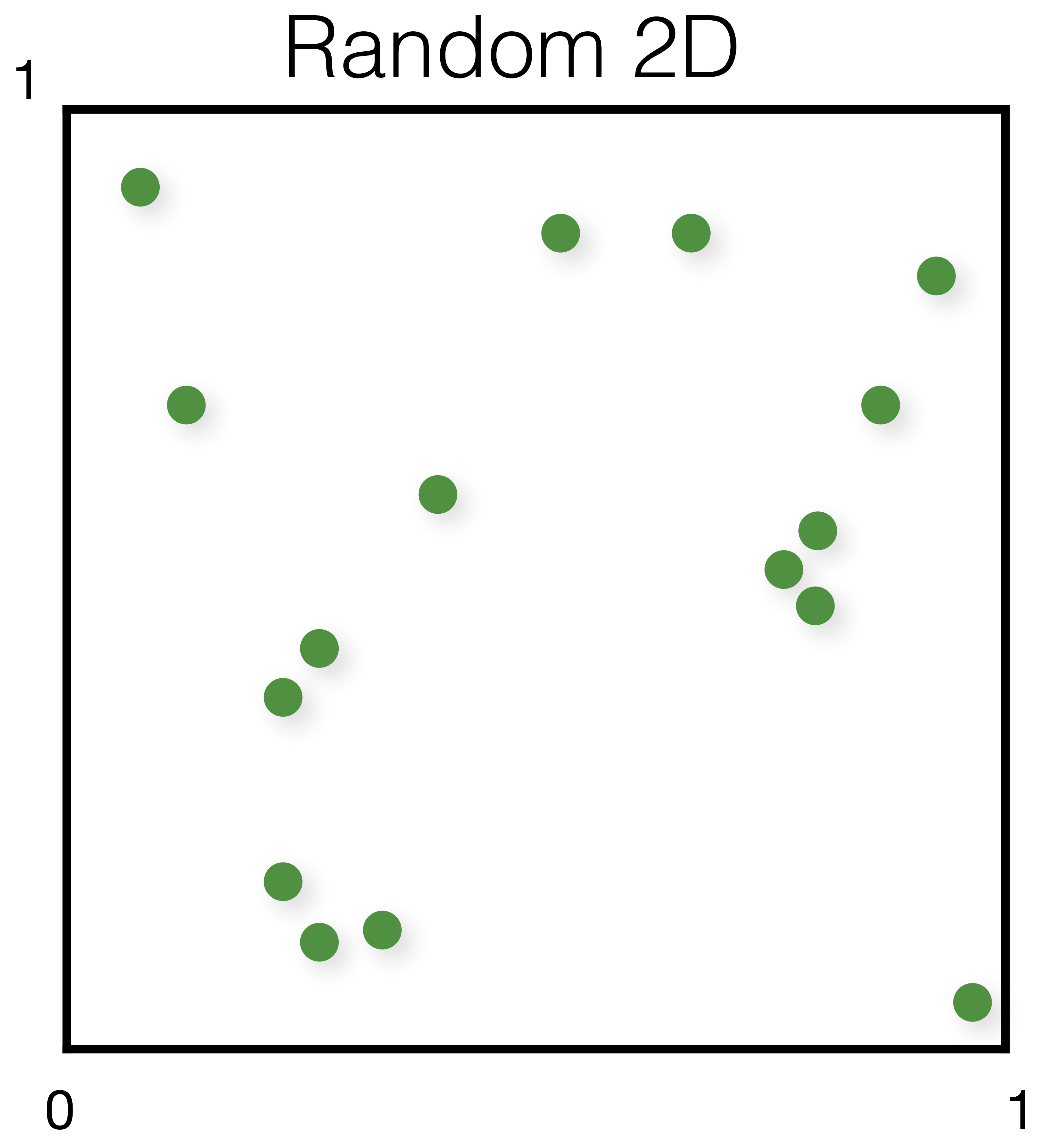
- Given the arbitrary interval $[a, b]$ in the domain, integrating the PDF gives the probability that a RV lies inside that interval:

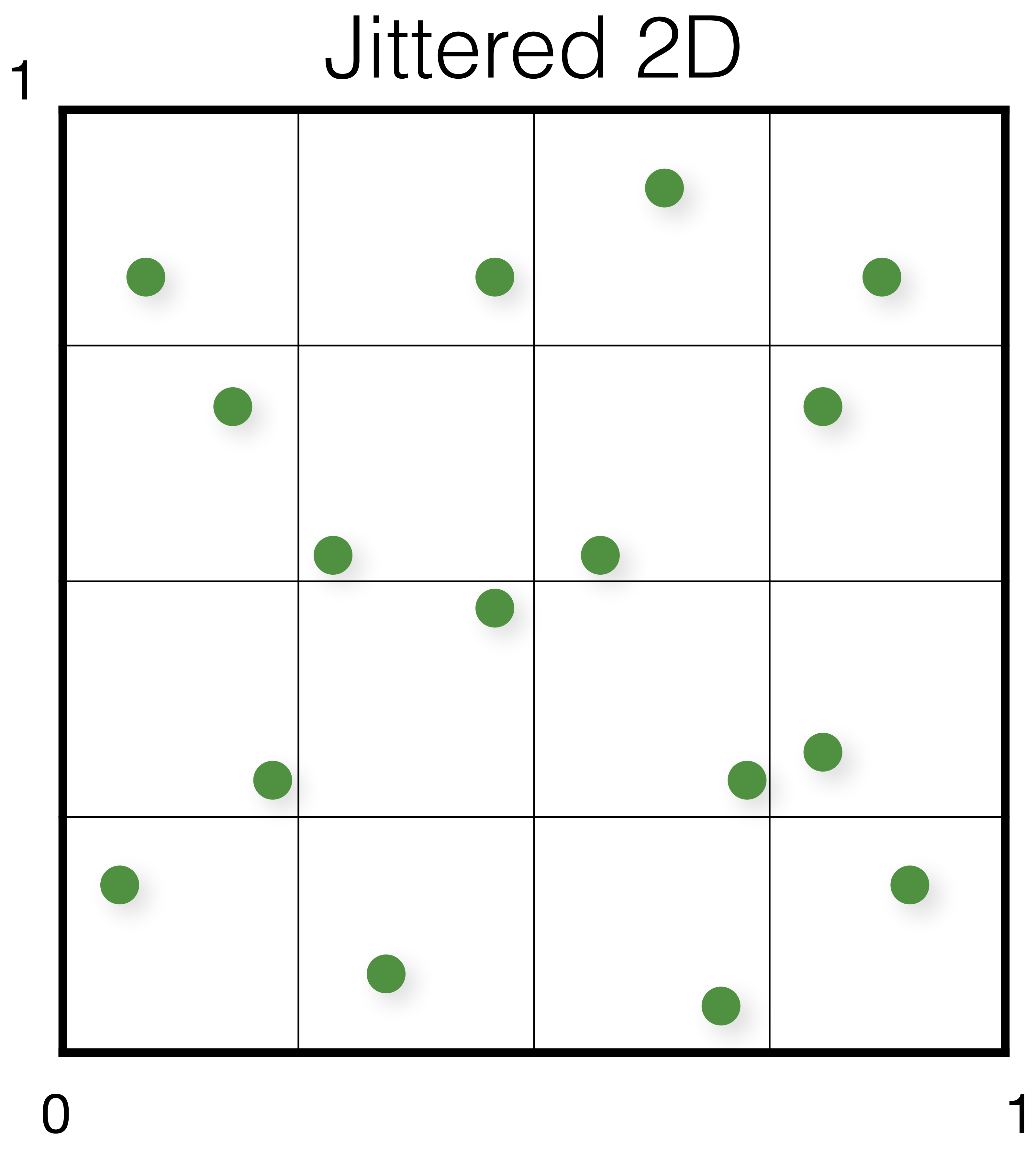
$$P(x \in [a, b]) = \int_a^b p(x)dx$$



Examples: Sampling PDFs

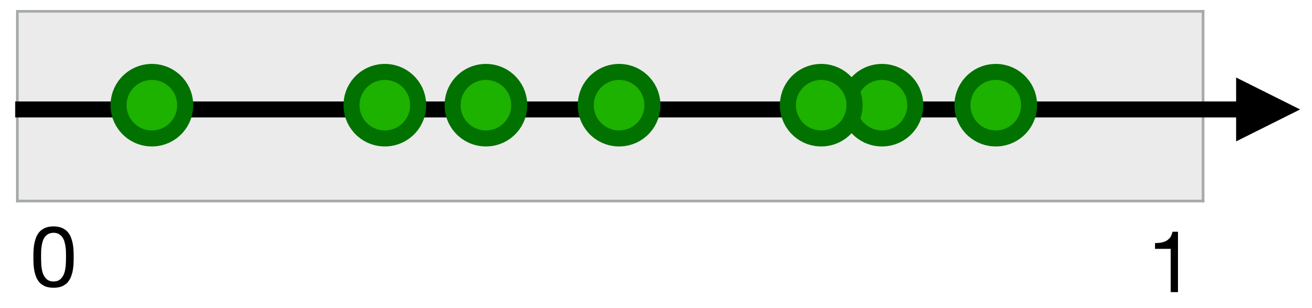
Constant Sampling PDFs





Constant Sampling PDFs

Random 1D

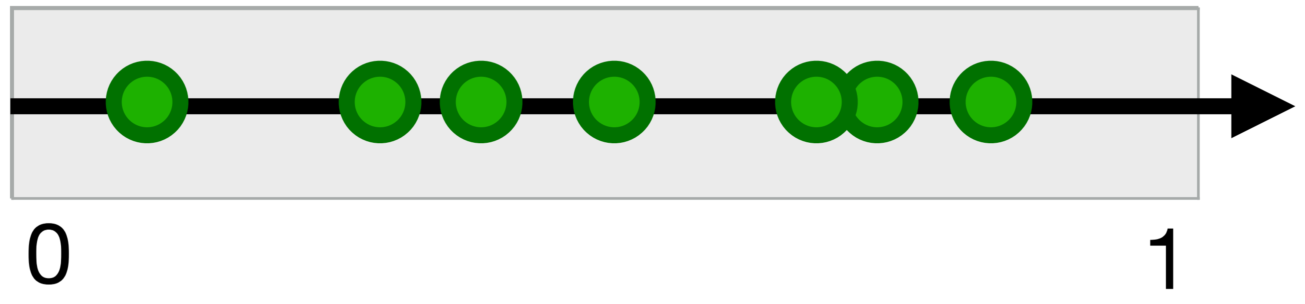


$$\xi \in [0, 1)$$

Sampling a unit domain with uniform random samples

Constant Sampling PDFs

Random 1D

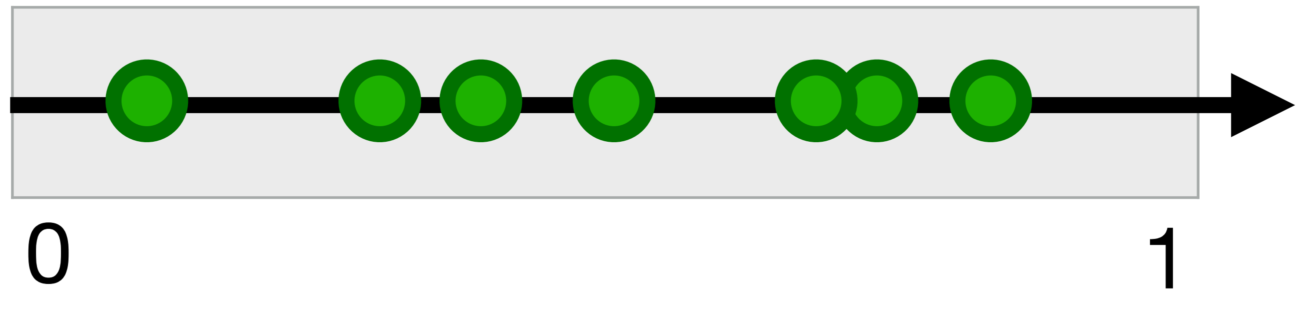


$$\xi \in [0, 1)$$

Sampling a unit domain with uniform random samples

Constant Sampling PDFs

Random 1D



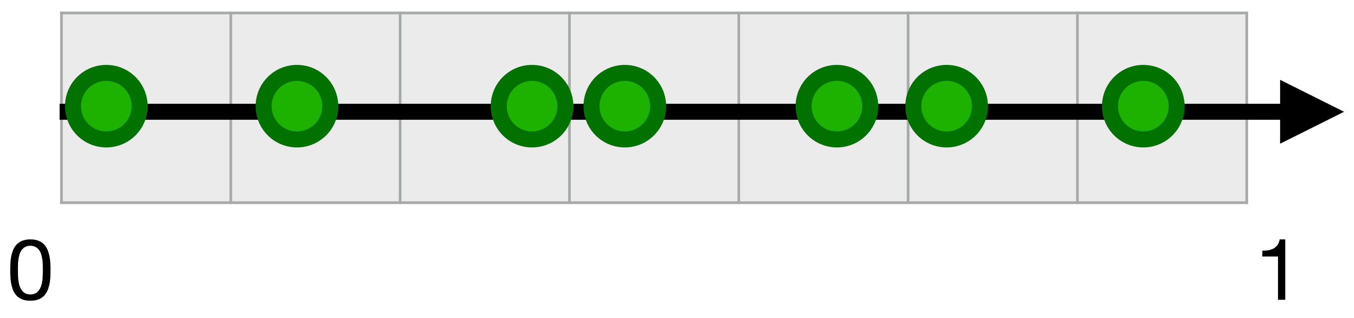
$$\xi \in [0,1)$$

$$p(x) = \begin{cases} \mathrm{C} & x \in [0,1) \\ 0 & \text{otherwise} \end{cases}$$

Sampling a unit domain with uniform random samples

Constant Sampling PDFs

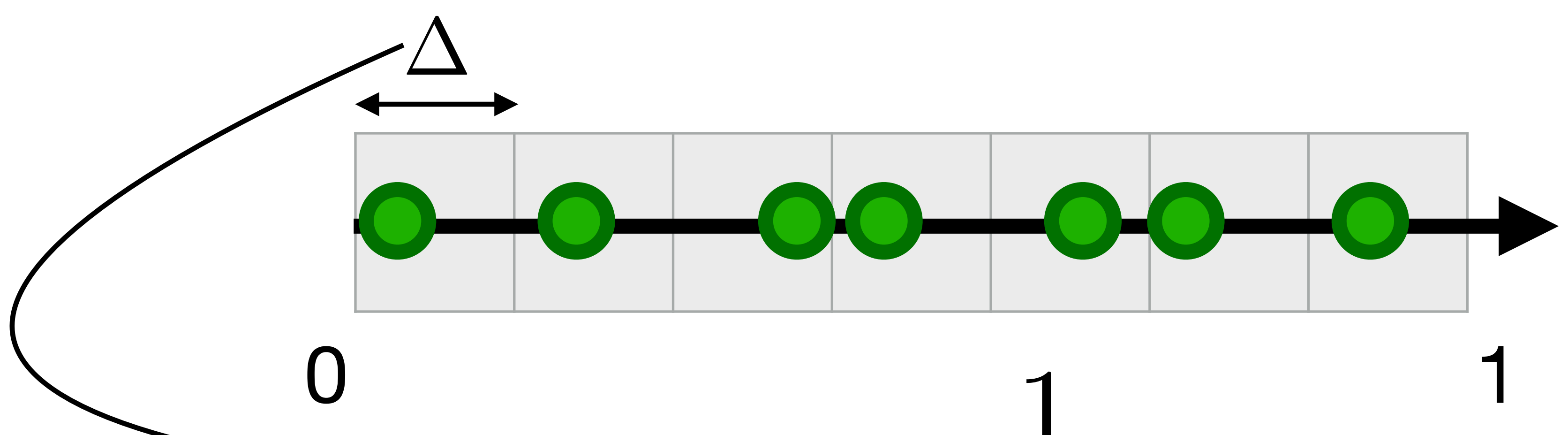
Jittered 1D



Sampling each stratum with uniform random samples

Constant Sampling PDFs

Jittered 1D

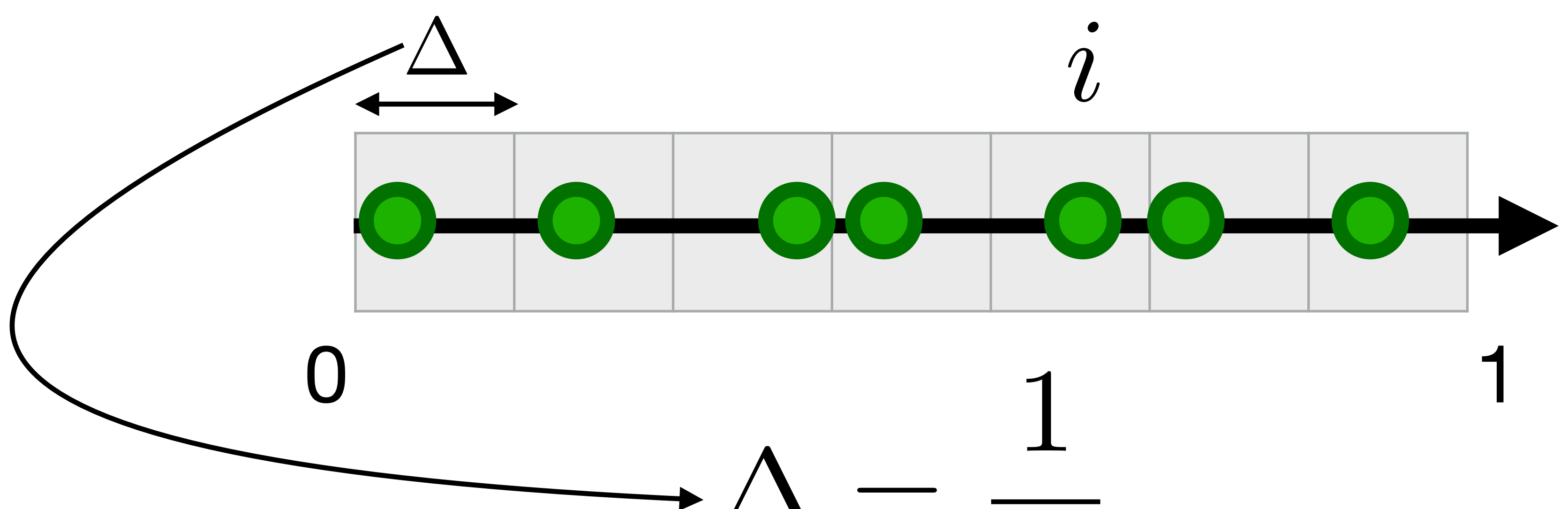


$$\Delta = \frac{1}{N}$$

Sampling each stratum with uniform random samples

Constant Sampling PDFs

Jittered 1D



$$\Delta = \frac{1}{N}$$

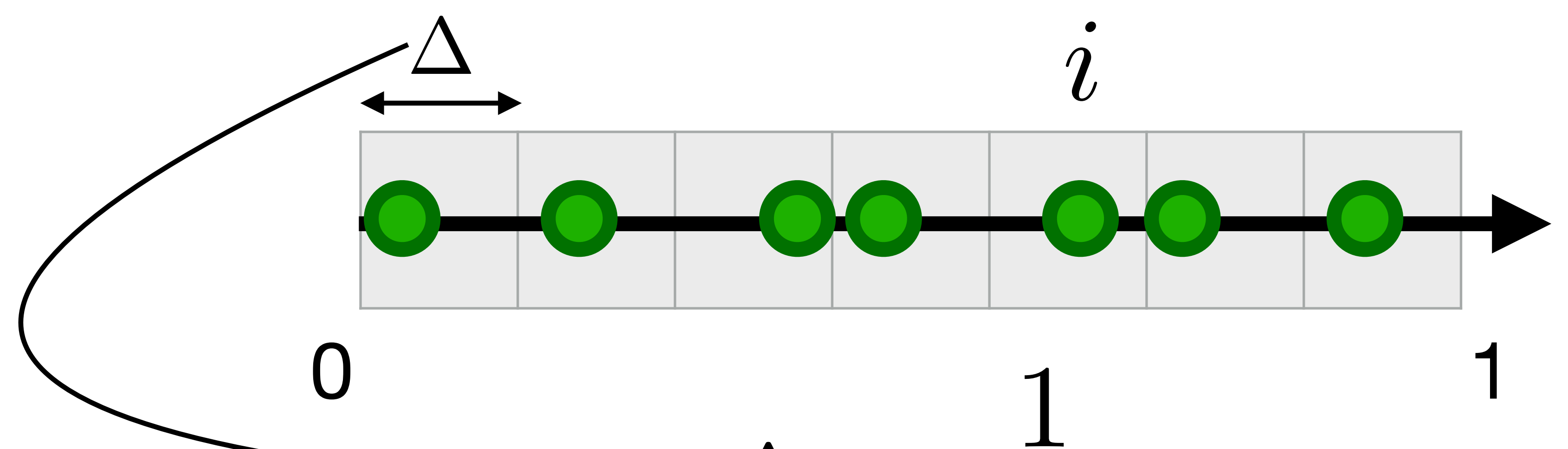
Probability density of generating a sample in an i-th stratum is given by:

$$p(x_i) = ???$$

Sampling each stratum with uniform random samples

Constant Sampling PDFs

Jittered 1D



$$\Delta = \frac{1}{N}$$

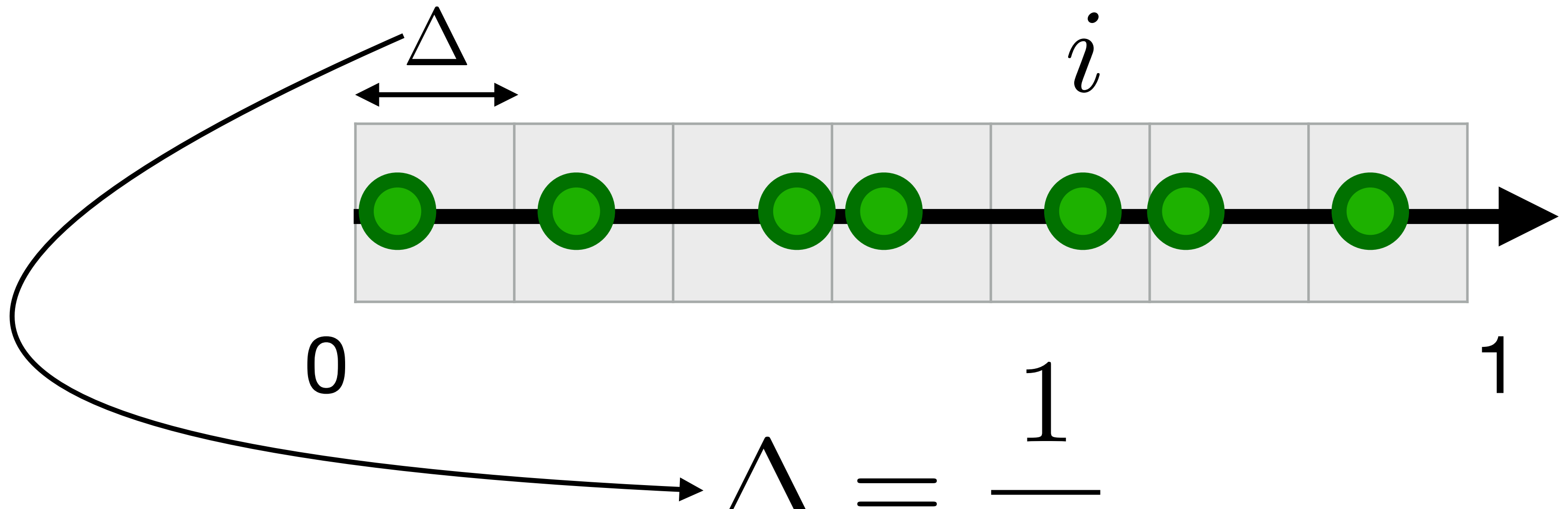
Probability density of generating a sample in an i-th stratum is given by:

$$p(x_i) = \begin{cases} N & x \in [\frac{i}{N}, \frac{i+1}{N}) \\ 0 & \text{otherwise} \end{cases}$$

Sampling each stratum with uniform random samples

Joint PDFs

Jittered 1D



$$\Delta = \frac{1}{N}$$

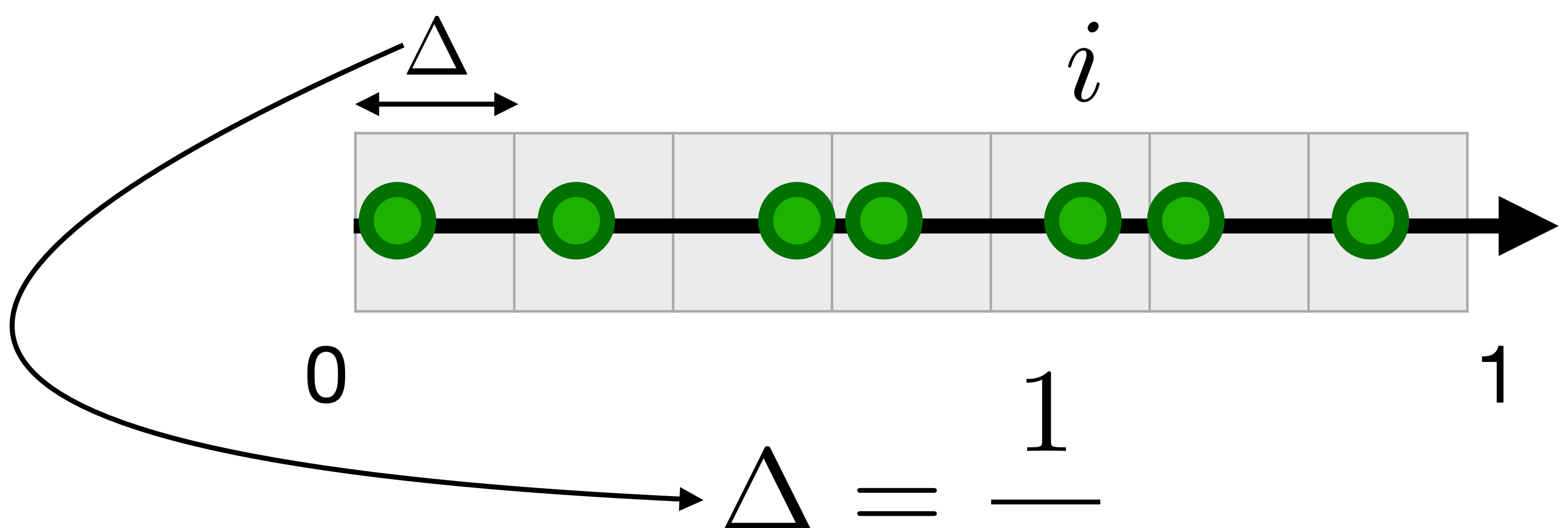
First, we divide the domain into equal strata.

Second, we sample the domain.

This implies that two samples are correlated to each other.

Joint PDFs

Jittered 1D



$$\Delta = \frac{1}{N}$$

First, we divide the domain into equal strata.

Second, we sample the domain.

This implies that two samples are correlated to each other.

For two different strata i and j, what is the joint PDF for jittered sampling ?

$$p(x_i, x_j) = ???$$

Conditional and Marginal PDFs

Joint PDF

For two random variables X_1 and X_2, the joint PDF $p(x_1, x_2)$ is given by:

Joint PDF

For two random variables X_1 and X_2, the joint PDF $p(x_1, x_2)$ is given by:

$$p(x_1, x_2) = p(x_2|x_1)p(x_1)$$

Joint PDF

For two random variables X_1 and X_2, the joint PDF $p(x_1, x_2)$ is given by:

$$p(x_1, x_2) = p(x_2|x_1)p(x_1)$$

where,

$X_1 = x_1$ $\quad$ $p(x_2|x_1)$: conditional density function

$X_2 = x_2$ $\quad$ $p(x_1)$: marginal density function

Joint PDF

For two random variables X_1 and X_2, the joint PDF $p(x_1, x_2)$ is given by:

$$p(x_1, x_2) = \underline{p(x_2|x_1)}\,\underline{p(x_1)}$$

where,

$X_1 = x_1$ $\boxed{p(x_2|x_1)}$: conditional density function

$X_2 = x_2$ $\boxed{p(x_1)}$: marginal density function

Joint PDF

For two random variables X_1 and X_2, the joint PDF $p(x_1, x_2)$ is given by:

$$p(x_1, x_2) = \underline{p(x_1|x_2)}\,\underline{p(x_2)}$$

where,

$$X_1 = x_1 \qquad \boxed{p(x_1|x_2)} \text{ : conditional density function}$$

$$X_2 = x_2 \qquad \boxed{p(x_2)} \text{ : marginal density function}$$

Marginal PDF

$$p(x_1) = \int_{\mathbb{R}} p(x_1, x_2) dx_2$$

$$p(x_2) = \int_{\mathbb{R}} p(x_1, x_2) dx_1$$

We integrate out one of the variable.

Conditional PDF

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)}$$

$$p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)}$$

The conditional density function is the density function for x_i given that some particular x_j has been chosen.

Conditional PDF

If both x_1 and x_2 are independent then:

$$p(x_1|x_2) = p(x_1)$$

$$p(x_2|x_1) = p(x_2)$$

Conditional PDF

If both x_1 and x_2 are independent then:

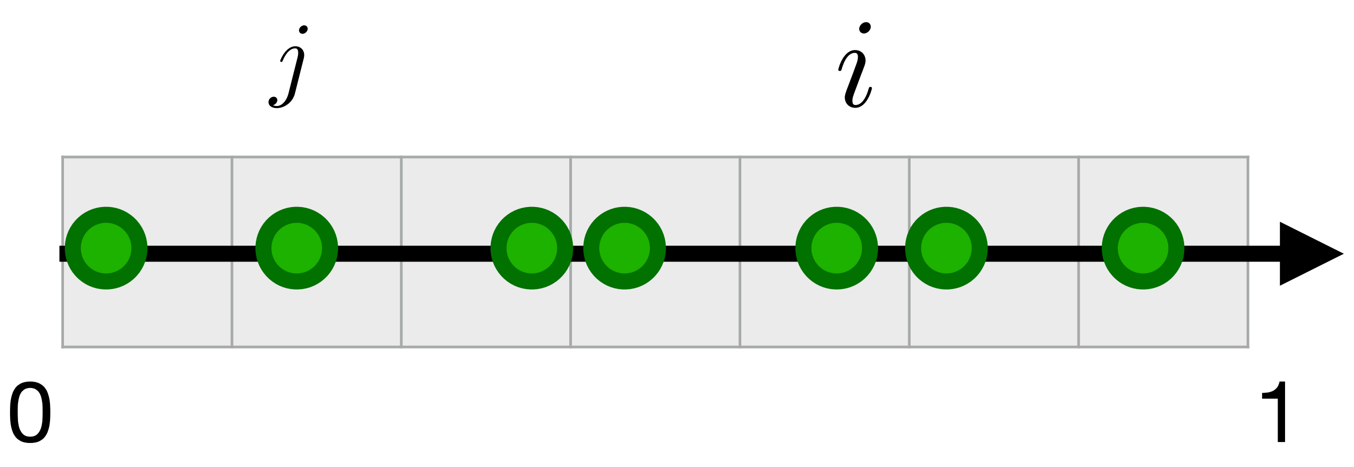
$$p(x_1|x_2) = p(x_1)$$

$$p(x_2|x_1) = p(x_2)$$

That gives:

$$p(x_1, x_2) = p(x_1)p(x_2)$$

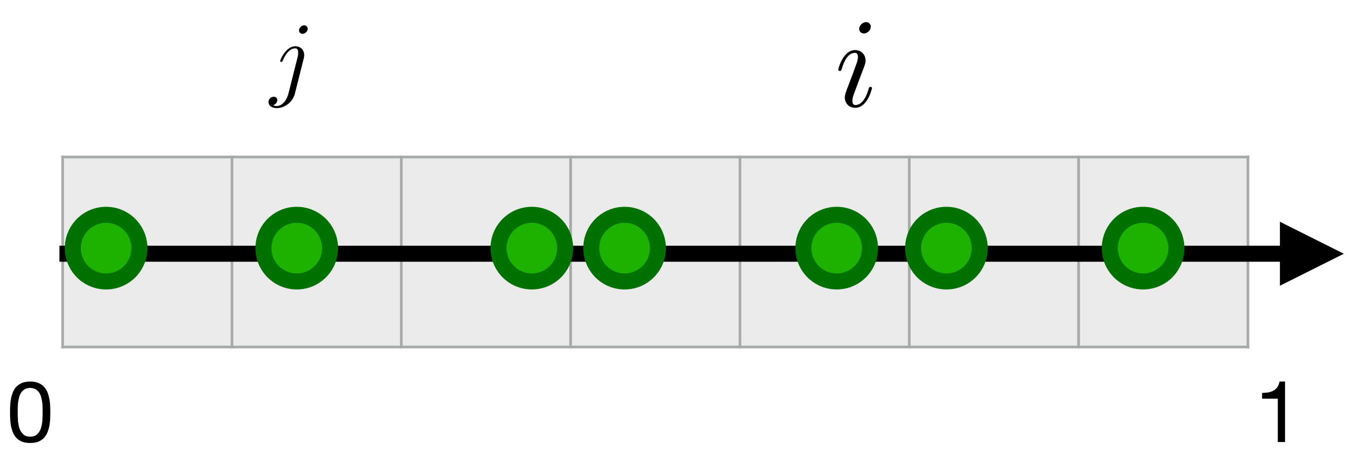
Joint PDF of Jittered 1D Sampling



For two different strata i and j, what is the joint PDF for jittered sampling ?

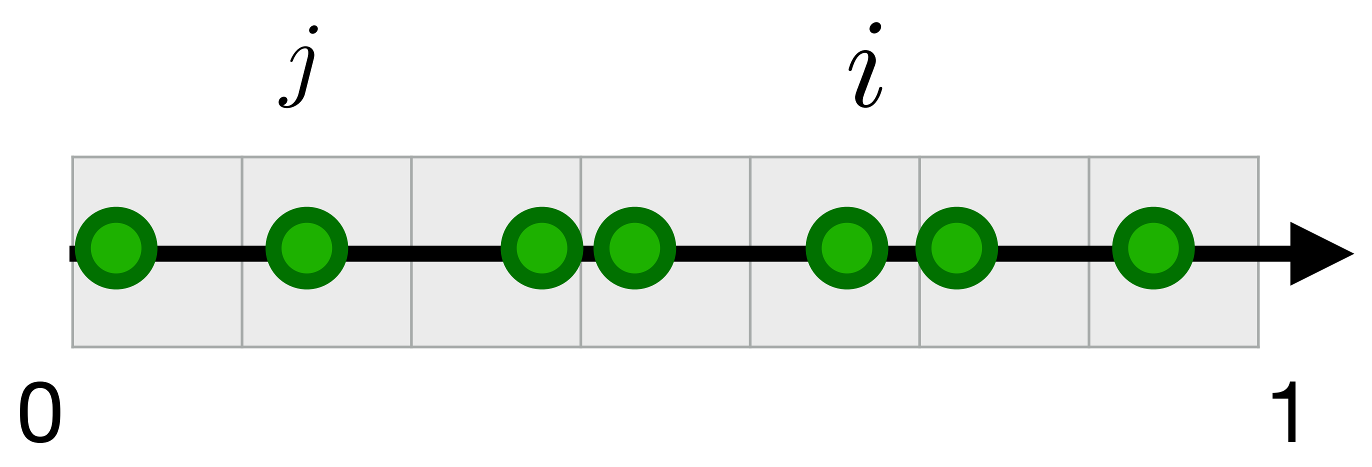
$$p(x_i, x_j) = ???$$

Joint PDF of Jittered 1D Sampling



$$p(x_1, x_2) = p(x_1|x_2)p(x_2)$$

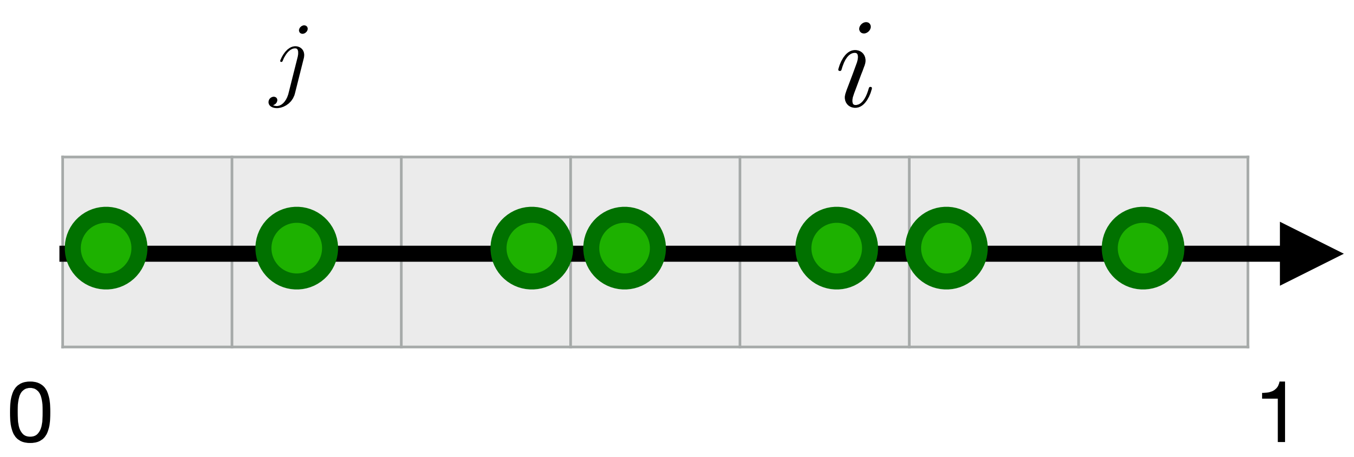
Joint PDF of Jittered 1D Sampling



$$p(x_1, x_2) = p(x_1|x_2)p(x_2)$$

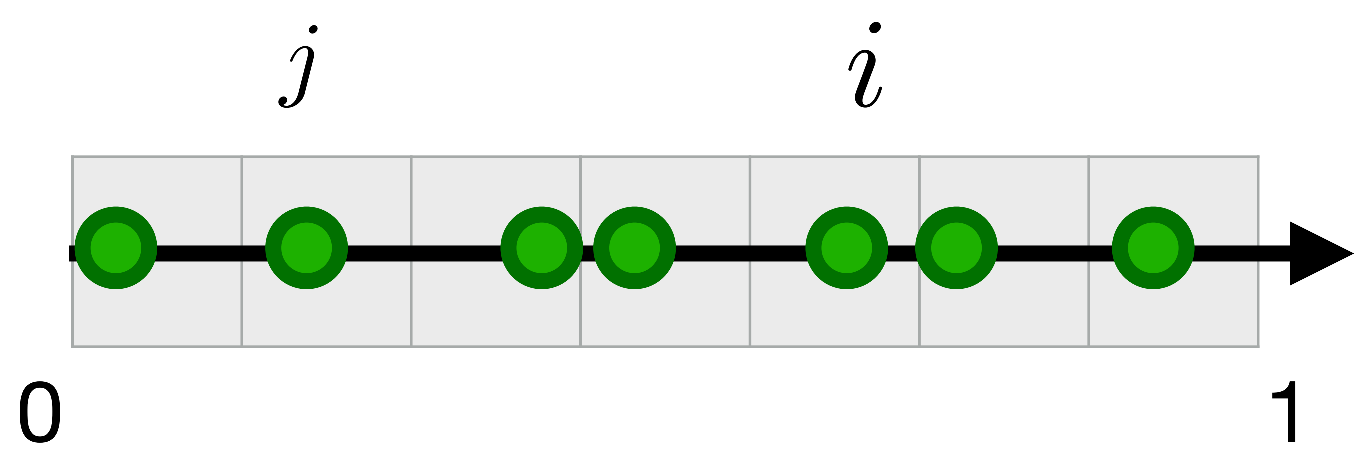
$$p(x_1, x_2) = p(x_1)p(x_2)$$

Joint PDF of Jittered 1D Sampling



$$p(x_i, x_j) = \begin{cases} p(x_i)p(x_j) & i \neq j \\ 0 & otherwise \end{cases}$$

Joint PDF of Jittered 1D Sampling



$$p(x_i, x_j) = \begin{cases} p(x_i)p(x_j) & i \neq j \\ 0 & otherwise \end{cases}$$

$$p(x_i, x_j) = \begin{cases} N^2 & i \neq j \\ 0 & otherwise \end{cases}$$

Since, $p(x_i) = N$

Visual Break

Image rendered using PBRT

Expected Value

Expected value

- Expected value: average value of the variable

$$E[X] = \sum_{i=1}^{N} x_i p_i$$

- example: rolling a die

$$E[X] =$$

Expected value

- Expected value: average value of the variable

$$E[X] = \sum_{i=1}^{N} x_i p_i$$

- example: rolling a die

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

Expected value

- Properties:

$$E[X+Y] = E[X] + E[Y]$$

Expected value

- Properties:

$$E[X + Y] = E[X] + E[Y]$$
$$E[X + c] = E[X] + c$$

Expected value

- Properties:

$$E[X+Y] = E[X] + E[Y]$$
$$E[X+c] = E[X] + c$$
$$E[cX] = cE[X]$$

Estimating expected values

- To estimate the expected value of a variable
  - choose a set of random *values* based on the probability
  - average their results

$$E[X] \approx \frac{1}{N} \sum_{i=1}^{N} x_i$$

- example: rolling a die
  - roll 3 times: $\{3, 1, 6\} \rightarrow E[\mathrm{x}] \approx$

Estimating expected values

- To estimate the expected value of a variable
  - choose a set of random *values* based on the probability
  - average their results

$$E[X] \approx \frac{1}{N} \sum_{i=1}^{N} x_i$$

- example: rolling a die
  - roll 3 times: $\{3, 1, 6\} \rightarrow E[\mathrm{x}] \approx (3 + 1 + 6)/3 = 3.33$

Estimating expected values

- To estimate the expected value of a variable
  - choose a set of random *values* based on the probability
  - average their results

$$E[X] \approx \frac{1}{N} \sum_{i=1}^{N} x_i$$

- example: rolling a die
  - roll 3 times: $\{3, 1, 6\} \rightarrow E[\mathrm{x}] \approx (3 + 1 + 6)/3 = 3.33$

Estimating expected values

- To estimate the expected value of a variable
  - choose a set of random *values* based on the probability
  - average their results

$$E[X] \approx \frac{1}{N}\sum_{i=1}^{N} x_i$$

- example: rolling a die
  - roll 3 times: $\{3, 1, 6\} \rightarrow E[\mathrm{x}] \approx (3+1+6)/3 = 3.33$
  - roll 9 times: $\{3, 1, 6, 2, 5, 3, 4, 6, 2\} \rightarrow E[\mathrm{x}] \approx$

Estimating expected values

- To estimate the expected value of a variable
  - choose a set of random *values* based on the probability
  - average their results

$$E[X] \approx \frac{1}{N}\sum_{i=1}^{N} x_i$$

- example: rolling a die
  - roll 3 times: $\{3, 1, 6\} \rightarrow E[\mathrm{x}] \approx (3 + 1 + 6)/3 = 3.33$
  - roll 9 times: $\{3, 1, 6, 2, 5, 3, 4, 6, 2\} \rightarrow E[\mathrm{x}] \approx$

Estimating expected values

- To estimate the expected value of a variable
  - choose a set of random *values* based on the probability
  - average their results

$$E[X] \approx \frac{1}{N} \sum_{i=1}^{N} x_i$$

- example: rolling a die
  - roll 3 times: $\{3, 1, 6\} \rightarrow E[\mathrm{x}] \approx (3 + 1 + 6)/3 = 3.33$
  - roll 9 times: $\{3, 1, 6, 2, 5, 3, 4, 6, 2\} \rightarrow E[\mathrm{x}] \approx 3.51$

Law of large numbers

- By taking *infinitely* many samples, the error between the estimate and the expected value is *statistically* zero
  - the estimate will converge to the right value

$$\text{probability}\left[E[x] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i\right] = 1$$

Variance

Variance

- Variance: how much different from the average

$$\sigma^2[X] = E[(X - E[X])^2]$$

Variance

- Variance: how much different from the average

$$
\begin{aligned}
\sigma^2[X] &= E[(X - E[X])^2] \\
&= E[X^2 + E[X]^2 - 2XE[X]]
\end{aligned}
$$

Variance

- Variance: how much different from the average

$$\begin{aligned}\sigma^2[X] &= E[(X - E[X])^2] \\ &= E[X^2 + E[X]^2 - 2XE[X]] \\ &= E[X^2] + E[E[X]^2] - 2E[X]E[E[X]]\end{aligned}$$

Variance

- Variance: how much different from the average

$$
\begin{aligned}
\sigma^2[X] &= E[(X - E[X])^2] \\
&= E[X^2 + E[X]^2 - 2XE[X]] \\
&= E[X^2] + E[E[X]^2] - 2E[X]E[E[X]]] \\
&= E[X^2] + E[X]^2 - 2E[X]^2
\end{aligned}
$$

Variance

- Variance: how much different from the average

$$
\begin{aligned}
\sigma^2[X] &= E[(X - E[X])^2] \\
&= E[X^2 + E[X]^2 - 2XE[X]] \\
&= E[X^2] + E[E[X]^2] - 2E[X]E[E[X]]] \\
&= E[X^2] + E[X]^2 - 2E[X]^2 \\
&= E[X^2] - E[X]^2
\end{aligned}
$$

Variance

- Variance: how much different from the average

$$
\begin{aligned}
\sigma^2[X] &= E[(X - E[X])^2] \\
&= E[X^2 + E[X]^2 - 2XE[X]] \\
&= E[X^2] + E[E[X]^2] - 2E[X]E[E[X]] \\
&= E[X^2] + E[X]^2 - 2E[X]^2 \\
&= E[X^2] - E[X]^2
\end{aligned}
$$

$$\boxed{\sigma^2[X] = E[X^2] - E[X]^2}$$

Variance

- example: Rolling a die
  - variance:

$$\boxed{\sigma^2[X] = E[X^2] - E[X]^2}$$

$$\sigma^2[X] = \ldots =$$

Variance

- example: Rolling a die
  - variance:

$$\boxed{\sigma^2[X] = E[X^2] - E[X]^2}$$

$$\sigma^2[X] = \ldots =$$

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

Variance

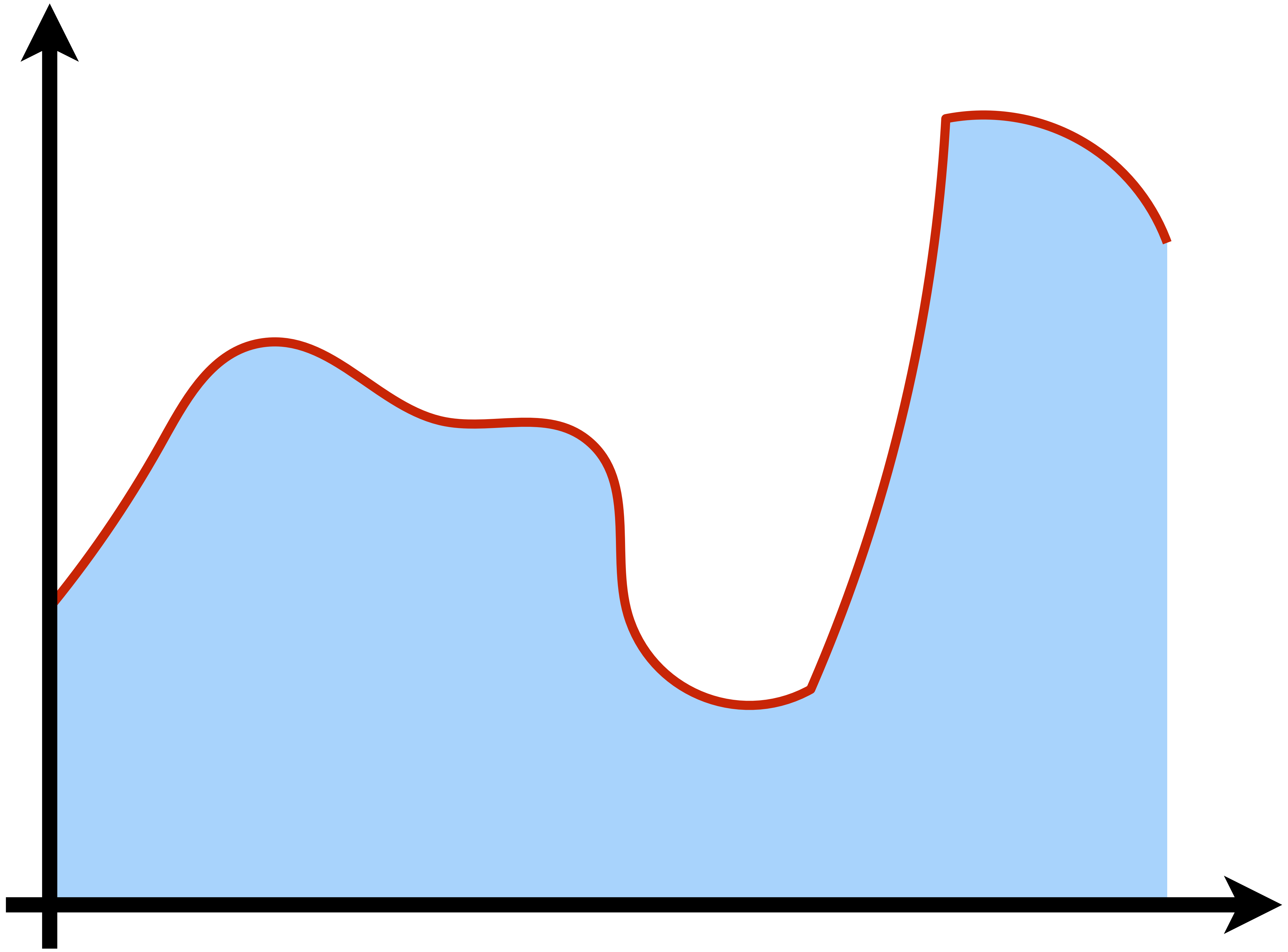
- example: Rolling a die
  - variance:

$$\boxed{\sigma^2[X] = E[X^2] - E[X]^2}$$

$$\sigma^2[X] = \ldots = 2.917$$

Monte Carlo Integration

$$I = \int_D f(x)\,\mathrm{d}x$$



Slide after Wojciech Jarosz

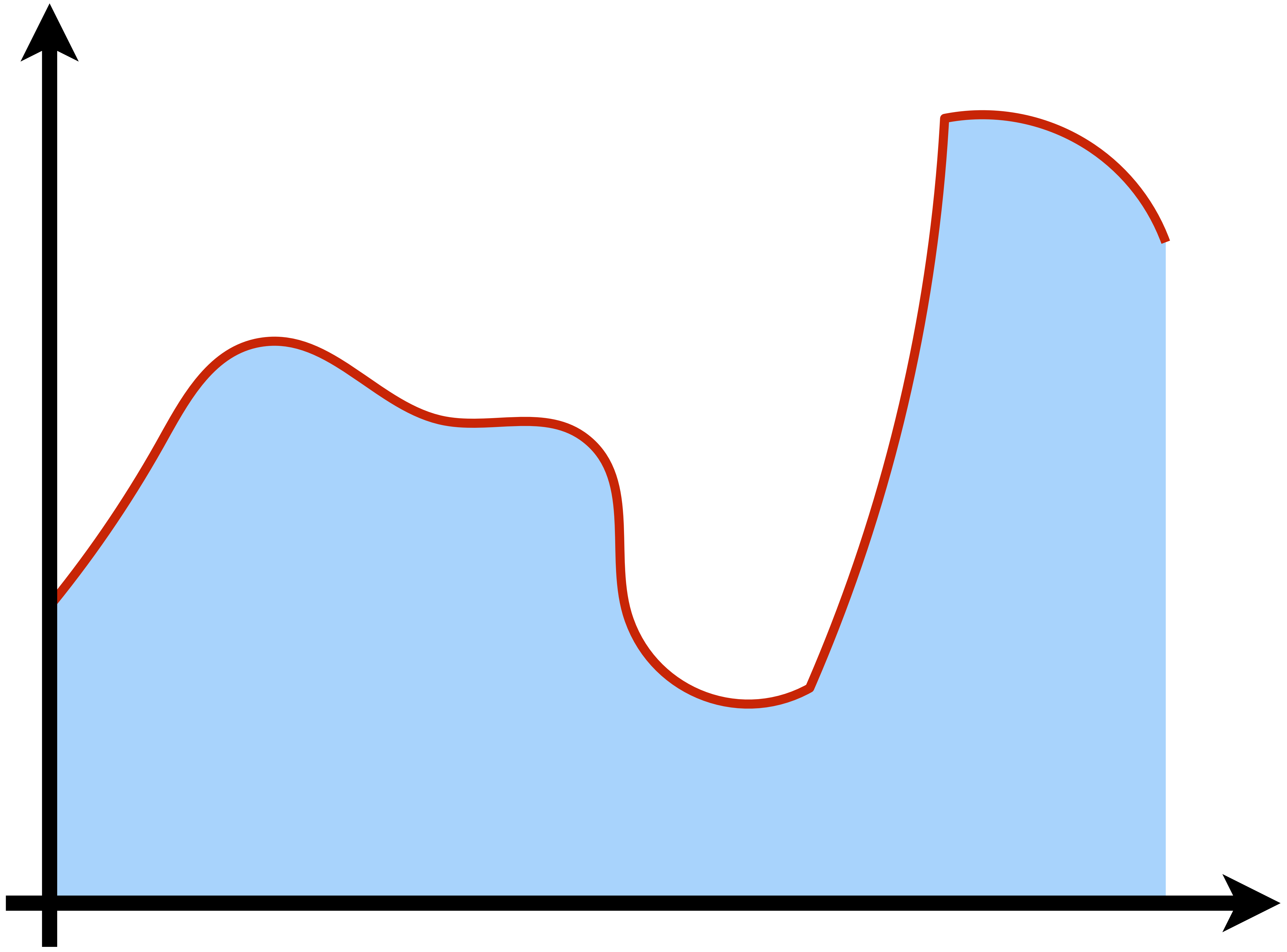
Questions ?

Image rendered using PBRT

Numerical Integration

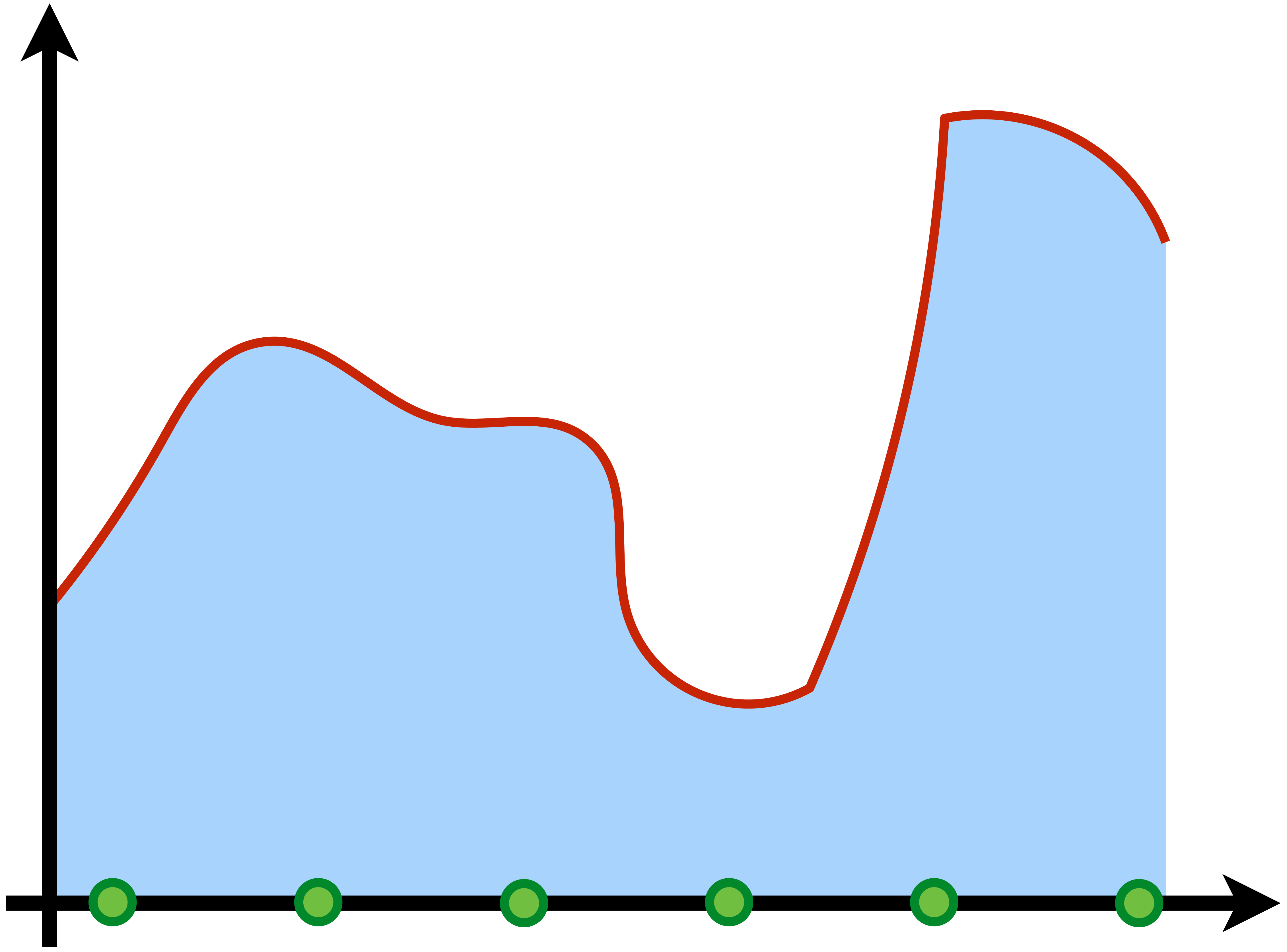
$$\int_a^b f(x)dx$$

- Analytic evaluation: accurate and fast



Numerical Integration

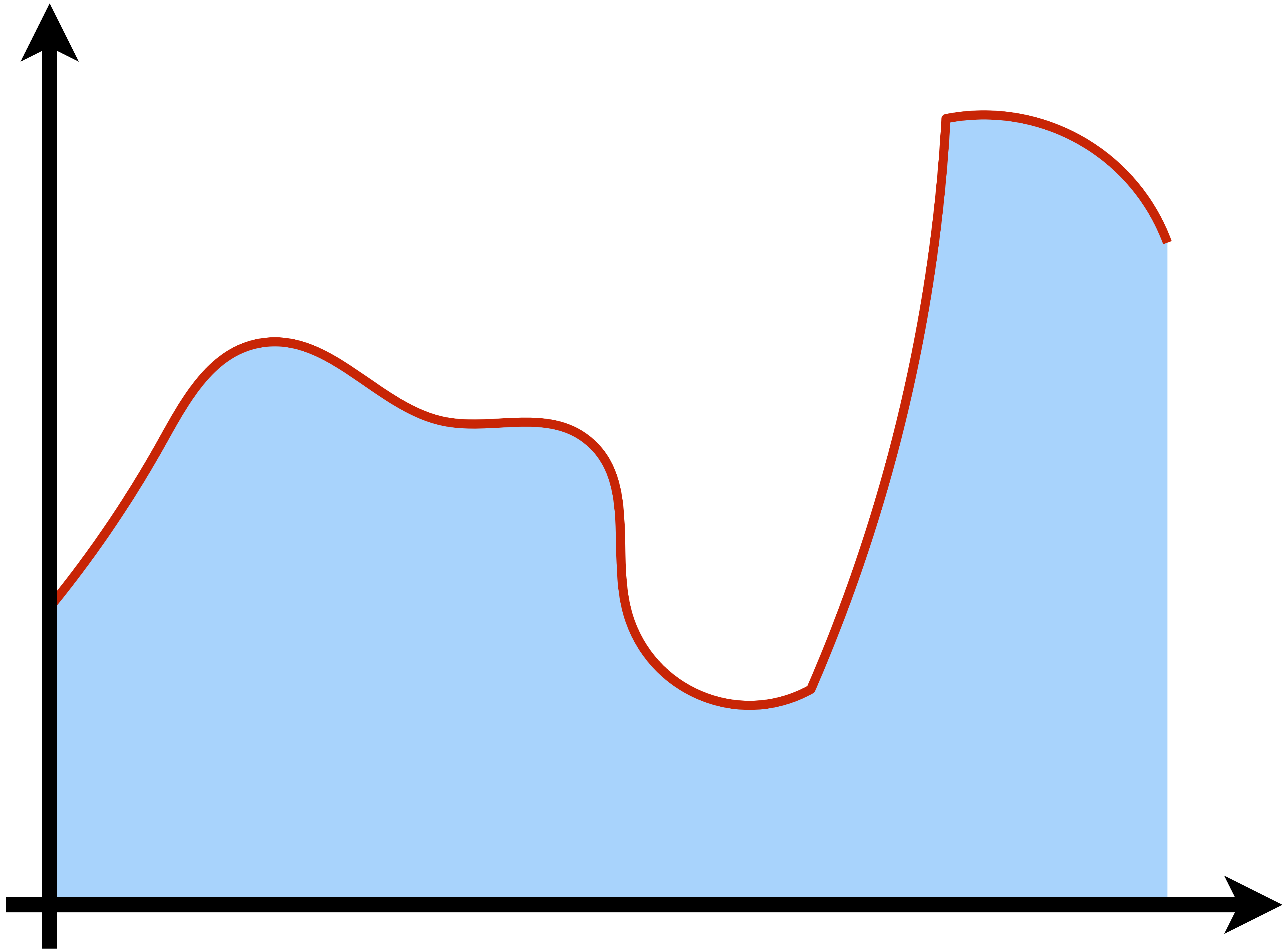
$$\int_a^b f(x)dx$$



- Numerical evaluations:
  - Provide only approximate solutions,
  - Rate of convergence is important
  - Often involves evaluations only at selected locations

Numerical Integration

$$\int_a^b f(x)dx$$



- Numerical quadrature: designed for 1D integrals
- Cubature/Quadratures: for higher dimensions

Numerical Integration

- Hybrid methods: First transform the integral analytically for simpler numerical handling

Numerical Integration

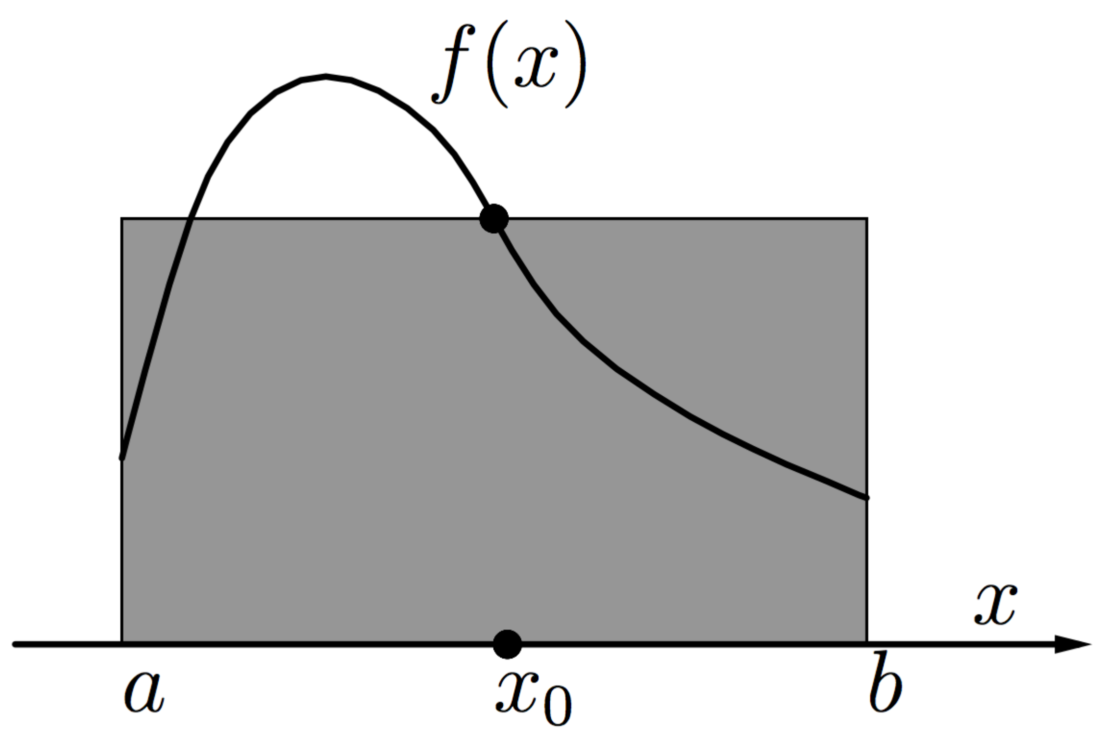
- A number of solutions are developed for the numeric solution of integrals
- Most prominent are the Quadrature rules, where the weights w_i and the sample positions x_i are determined in advance

$$\int_a^b f(x)dx = \sum_{i=1}^{N} w_i f(x_i)$$

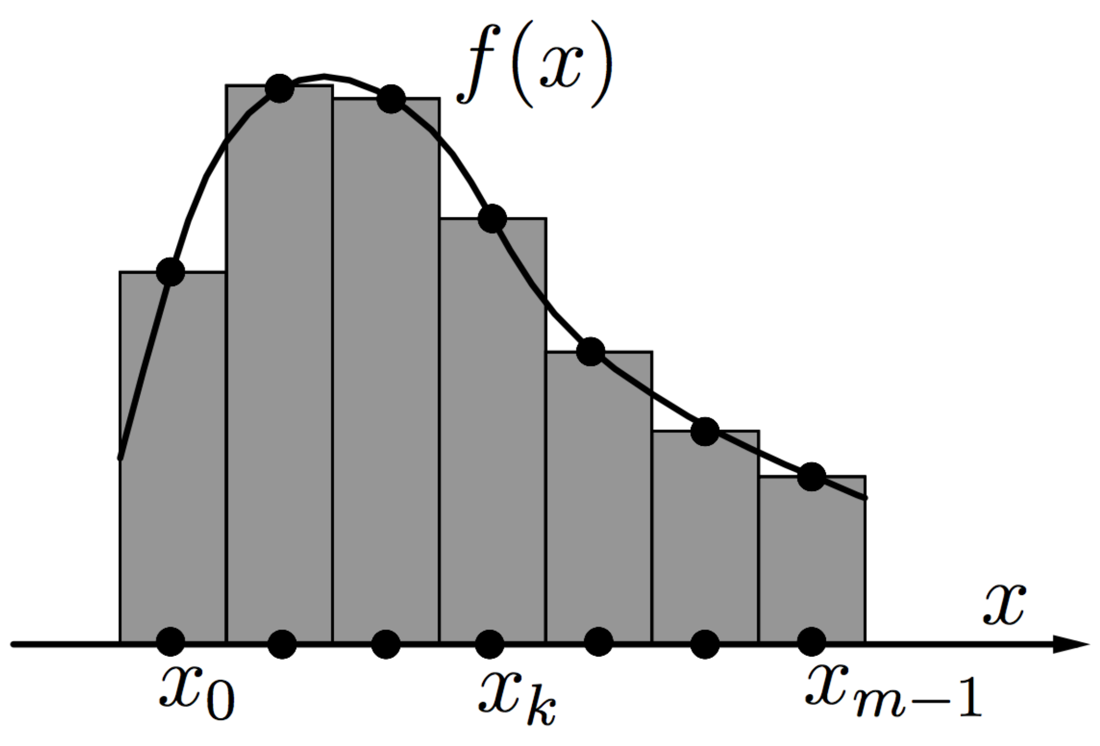
Quadrature rules

- Newton-Cots formula:
  - Midpoint rule (1 sample), Trapezoid rule (2 samples), Simpson rule (3 samples)...

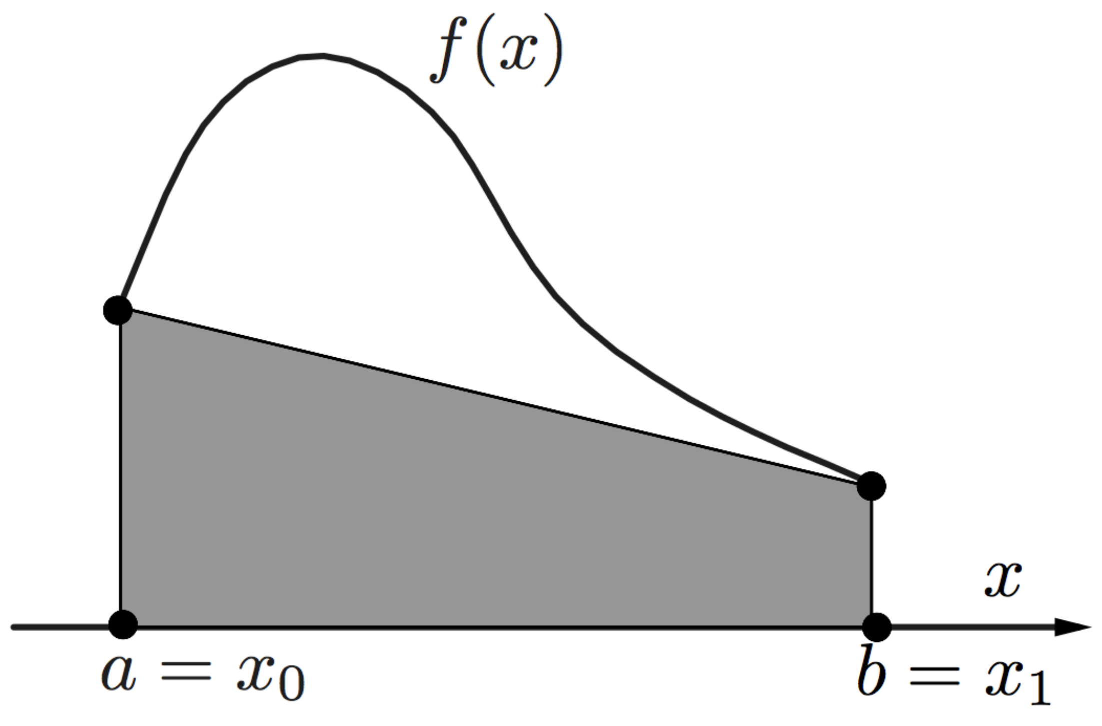
Image courtesy



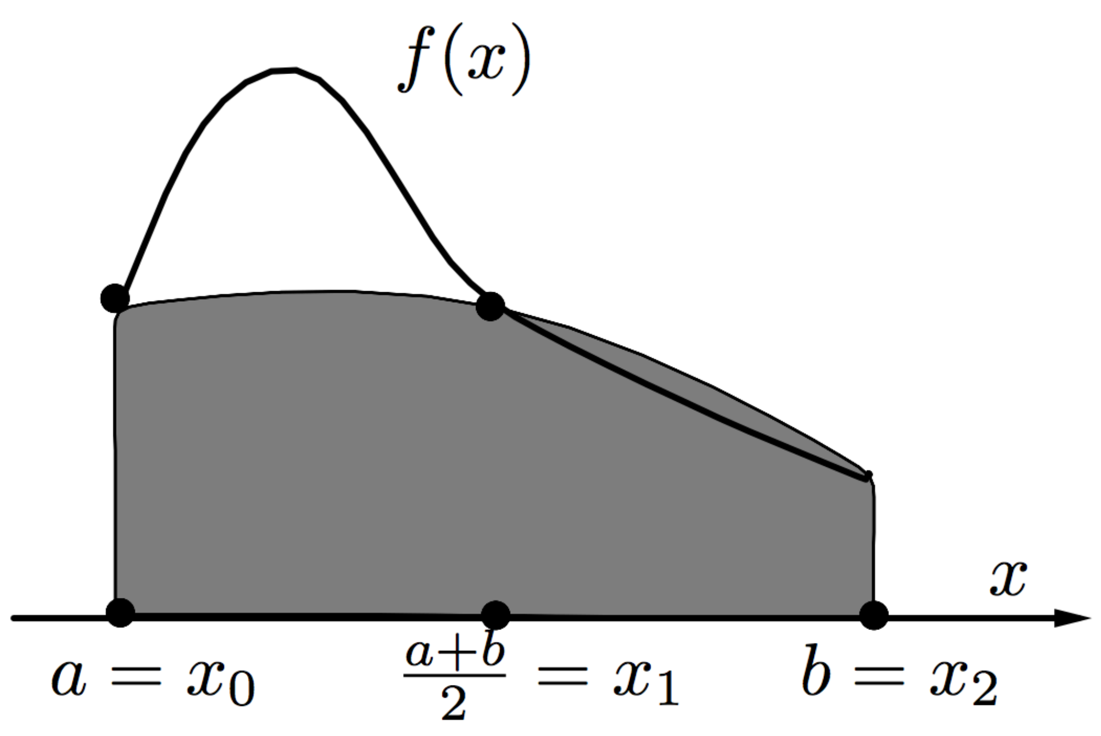
Midpoint formula



Composite midpoint formula



Trapezoidal formula



Cavalieri-Simpson formula

Quadrature rules

- Newton-Cots formula:
    - Midpoint rule (1 sample), Trapezoid rule (2 samples), Simpson rule (3 samples)...
        - Samples are nesting (for powers of 2)
        - Approximates the integral as sum of weighted, equidistant samples

Quadrature rules

- Gauss quadratures:
    - An n-point Gauss quadrature is constructed to yield exact results for polynomials of degree 2n-1 or less.
    - Extends freedom by allowing choice of sample locations
    - It doesn't nest (but nesting alternatives do exist)

Quadrature rules

- Gauss quadratures:
    - An n-point Gauss quadrature is constructed to yield exact results for polynomials of degree 2n-1 or less.
    - Extends freedom by allowing choice of sample locations
    - It doesn't nest (but nesting alternatives do exist)

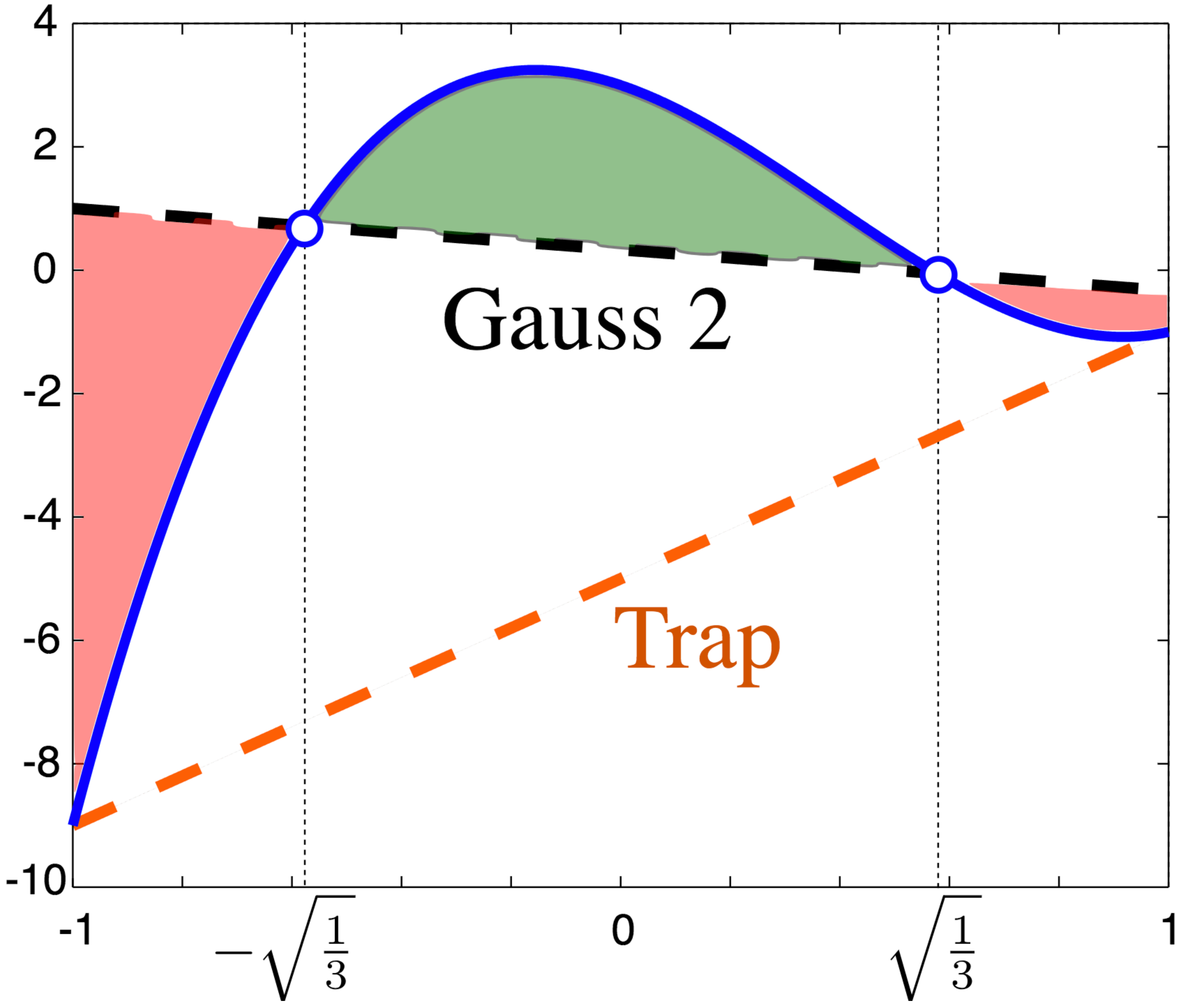


Image from Wikipedia

Quadrature rules

Newton-Cots formula*

Gauss quadratures*

Both approaches achieve convergence of the order $\mathcal{O}(N^{-r})$, given N samples and a smooth integrand that has r-continuous derivatives

***Interested students may refer to <u>this</u> link for more details.**

Numerical Integration: sD case

$$\int_a^b \dots \int_a^b f(x_1, ..., x_s) dx_1 ... dx_s = \sum_{i_1=1}^{N} \dots \sum_{i_s=1}^{N} w_{i_1} ... w_{i_s} f(x_{i_1}, ..., x_{i_s})$$

- Curse of dimensionality: requires N^s samples for s-dimensional integral
- Convergence drops to $\mathcal{O}(N^{-r/s})$
- Rules must be adapted to non-square domains (typical in rendering)

Monte Carlo Integration

- Independent of the dimensions
- Independent of the underlying topology of the domain
- Variance converges at $O(N^{-1})$ irrespective of the dimensions (N is the sample count)

Integral as Expected Value

$$\int_{Q^s} f(x) d\mu_s(x) = \int_{[0,1)^s} f(x) dx = \int_{[0,1)^s} \frac{f(x)}{p(x)} p(x) dx$$

$p(x)$: is an arbitrary probability density function over the domain

Integral as Expected Value

$$\int_{[0,1)^s} f(x)dx = \int_{[0,1)^s} \frac{f(x)}{p(x)} p(x) dx$$

$p(x)$: is an arbitrary probability density function over the domain

Integral as Expected Value

$$\int_{[0,1)^s} f(x)dx = \int_{[0,1)^s} \frac{f(x)}{p(x)} p(x)dx$$

$p(x)$: is an arbitrary probability density function over the domain

$$= \int_{[0,1)^s} \Big(\frac{f(x)}{p(x)}\Big) p(x)dx$$

$$= E\Big[\frac{f(x)}{p(x)}\Big]$$

$$\boxed{E[g(x)] = \int_Q g(x)p(x)dx}$$

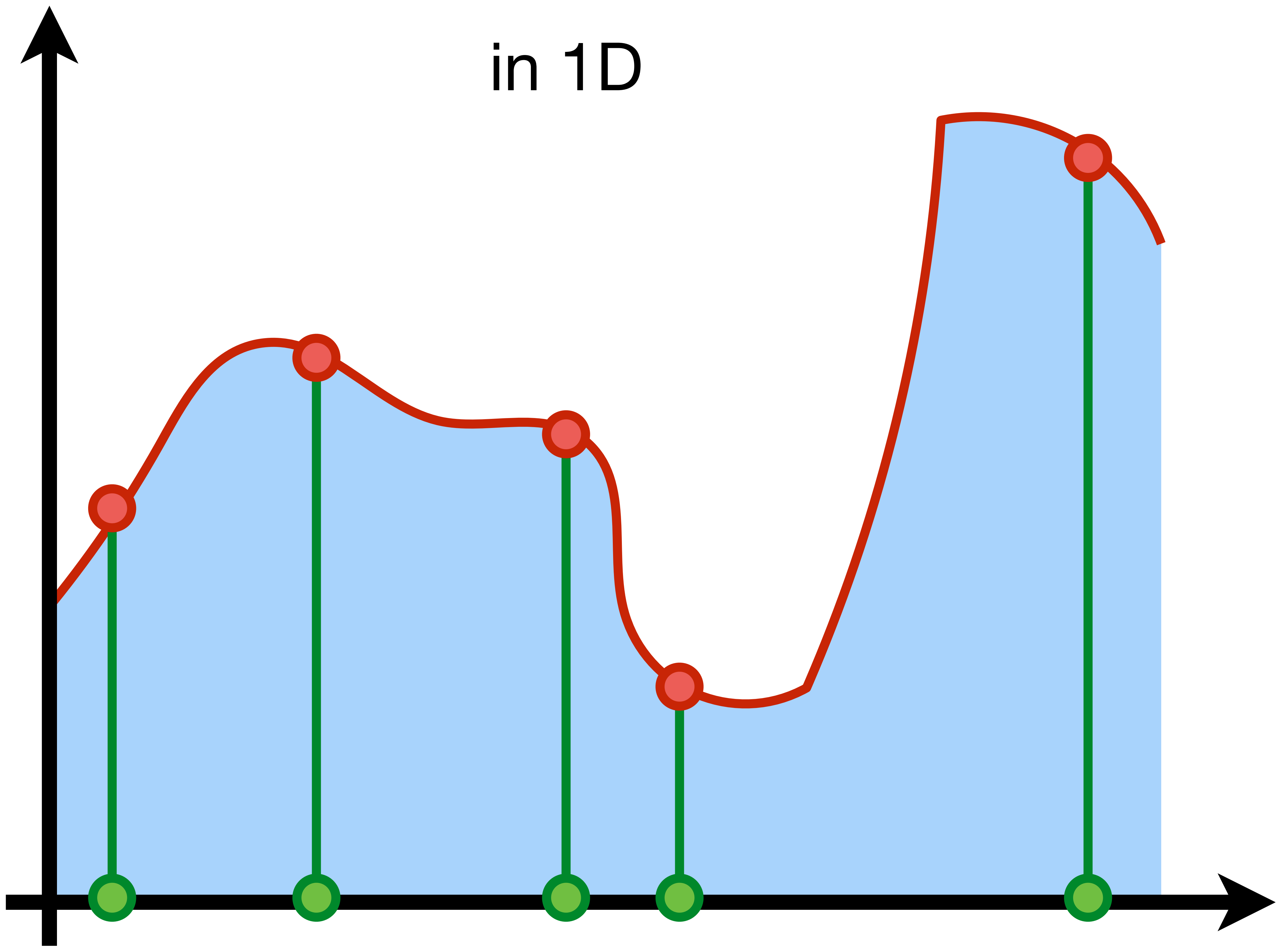
Integral as Expected Value

$$\int_{[0,1)^s} f(x)dx = E\left[\frac{f(x)}{p(x)}\right]$$

We are interested in the numerical computation of this expected value, leading to the highly important concept of **Monte Carlo Estimator**

Monte Carlo Estimator

$$\mathbf{I} = \int_0^1 f(x)dx$$

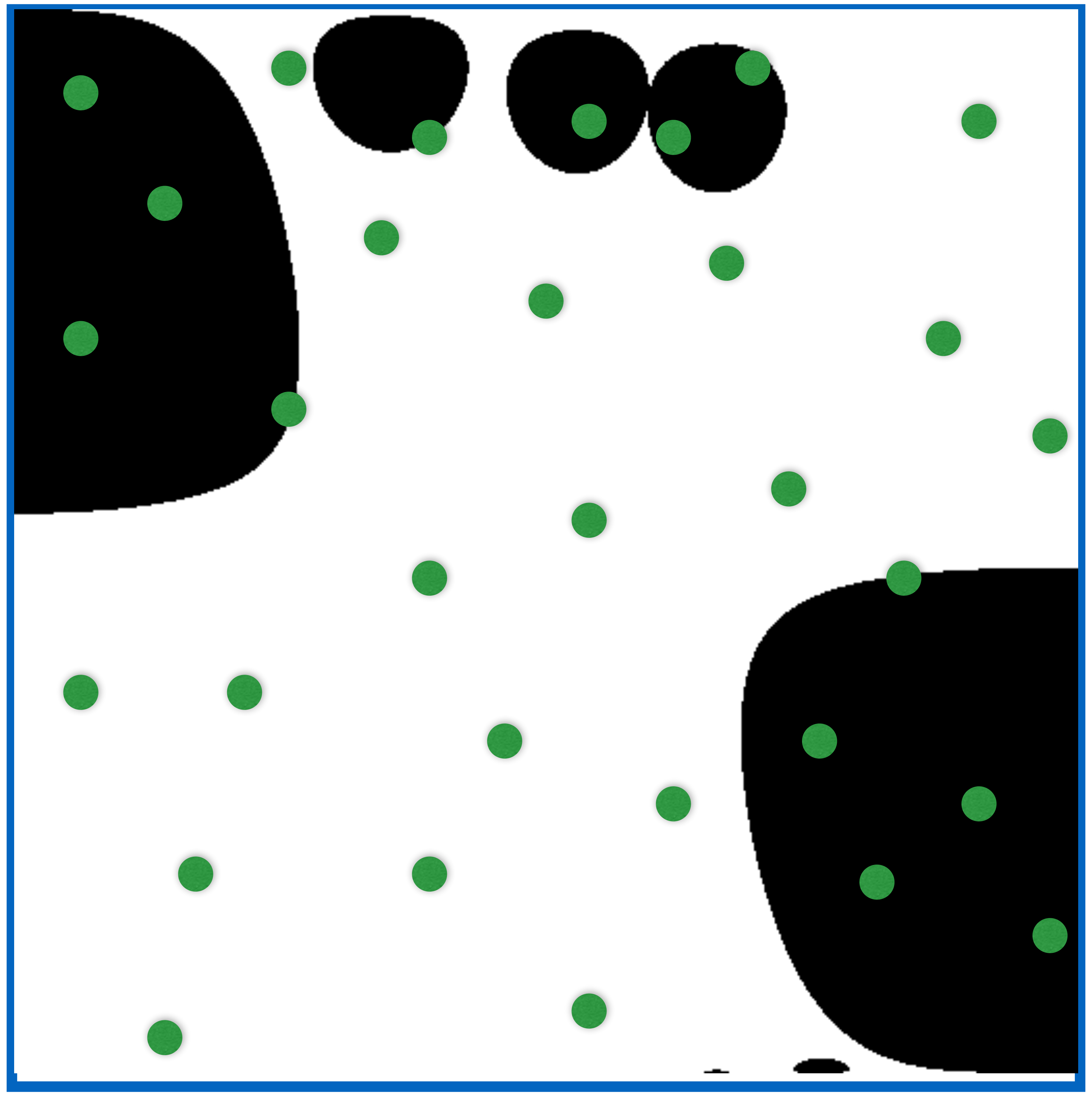


$p(x)$: is the probability density function from which samples are drawn

Monte Carlo Estimator

$$\mathbf{I} = \int_0^1 f(x)dx \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$$

in 2D

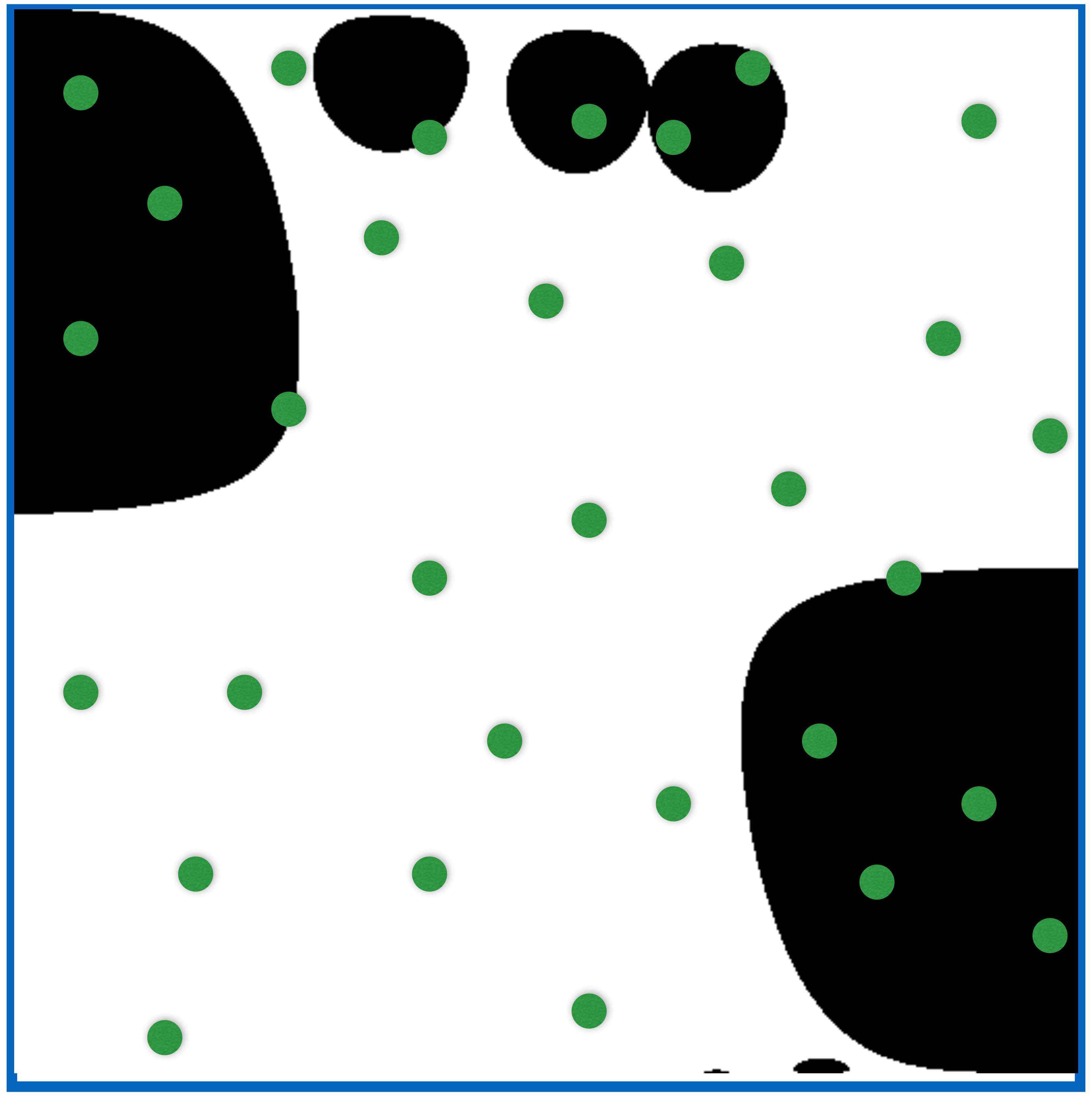


$p(x)$: is the probability density function from which samples are drawn

Monte Carlo Estimator

$$\mathbf{I} = \int_0^1 f(x)dx \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$$

in 2D



$p(x)$: is the probability density function from which samples are drawn

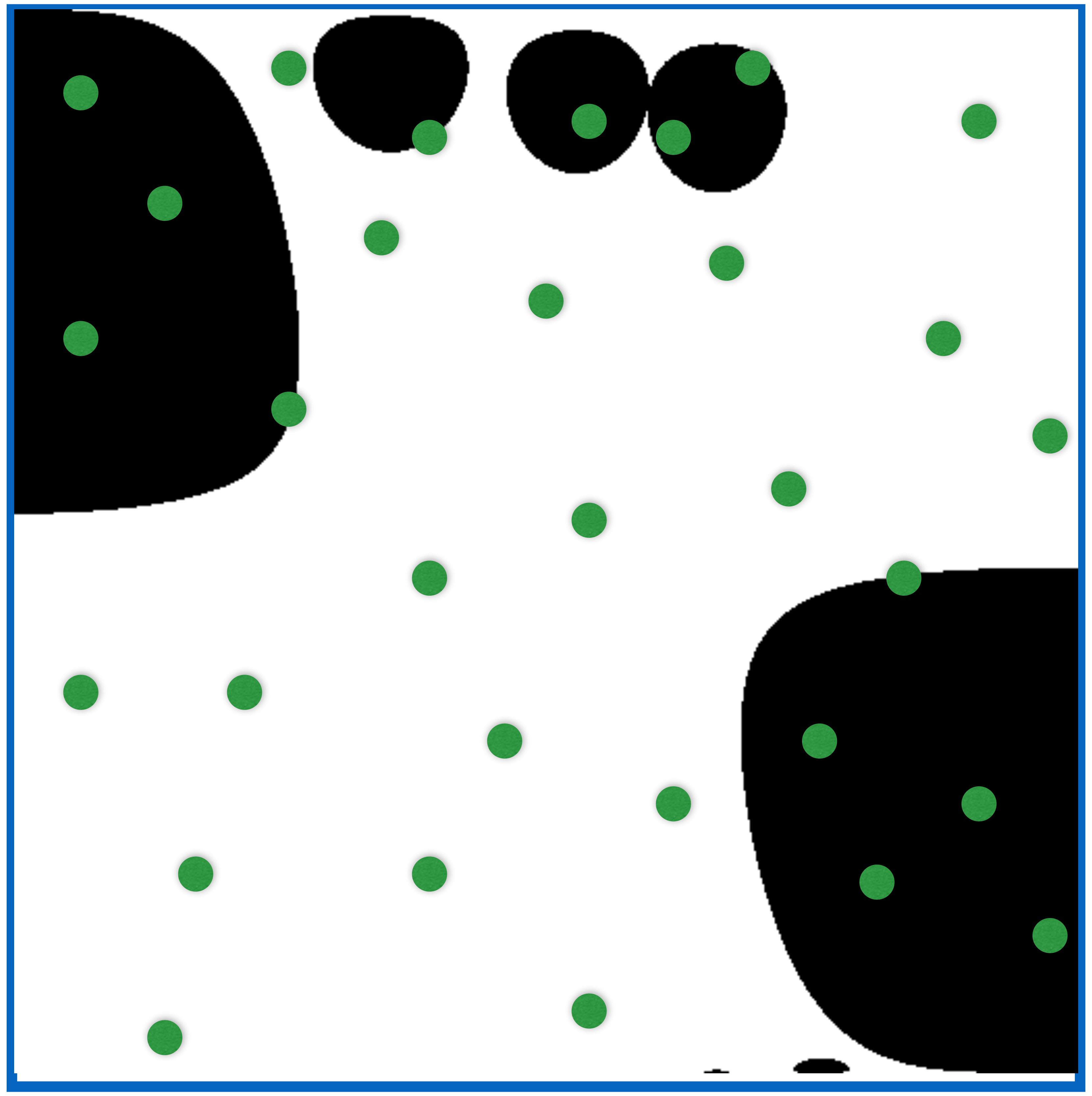
Monte Carlo Estimator

Secondary Estimator: $$\mathbf{I}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$$

$$\mathbf{I}_N = \frac{1}{N}\sum_{i=1}^{N}\mathbf{I}_1^i$$

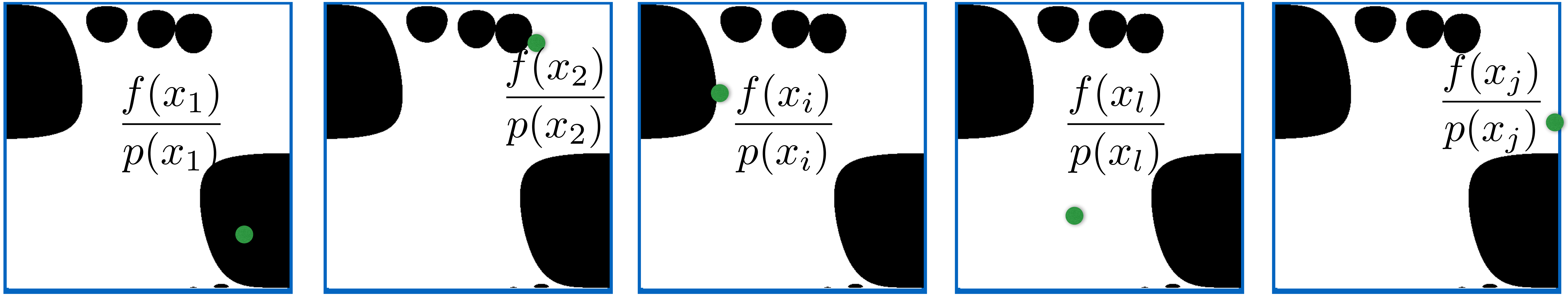
Primary Estimator: $$\mathbf{I}_1^i = \frac{f(x_i)}{p(x_i)}$$

$p(x)$: is the probability density function from which samples are drawn



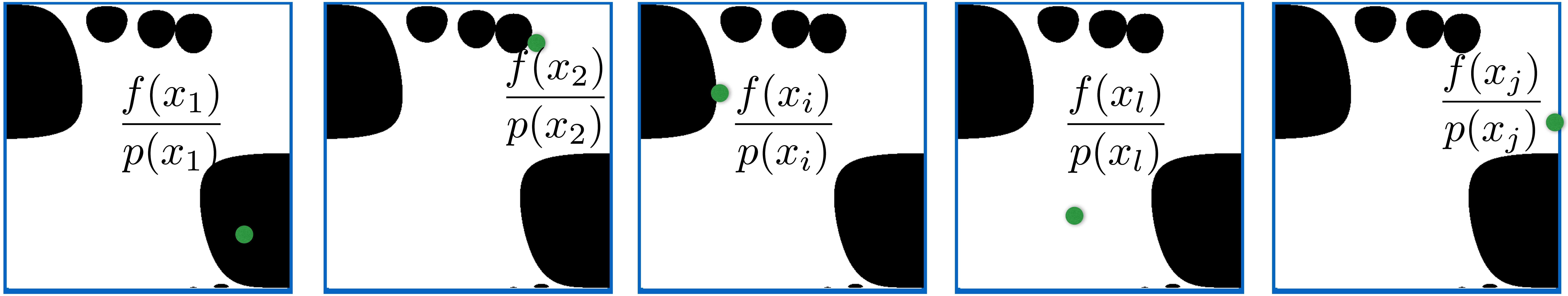
Monte Carlo Estimator

Primary Estimator: $\mathbf{I}_1^i = \frac{f(x_i)}{p(x_i)}$



Monte Carlo Estimator

Primary Estimator: $\mathbf{I}_1^i = \dfrac{f(x_i)}{p(x_i)}$



Monte Carlo Estimator

Primary Estimator: $$\mathbf{I}_1^i = \frac{f(x_i)}{p(x_i)}$$

$$\boldsymbol{\frac{f(x_1)}{p(x_1)}}$$

Monte Carlo Estimator

Primary Estimator: $$\mathbf{I}_1^i = \frac{f(x_i)}{p(x_i)}$$

$$\mathbf{I}_N = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)}$$

Monte Carlo Estimator

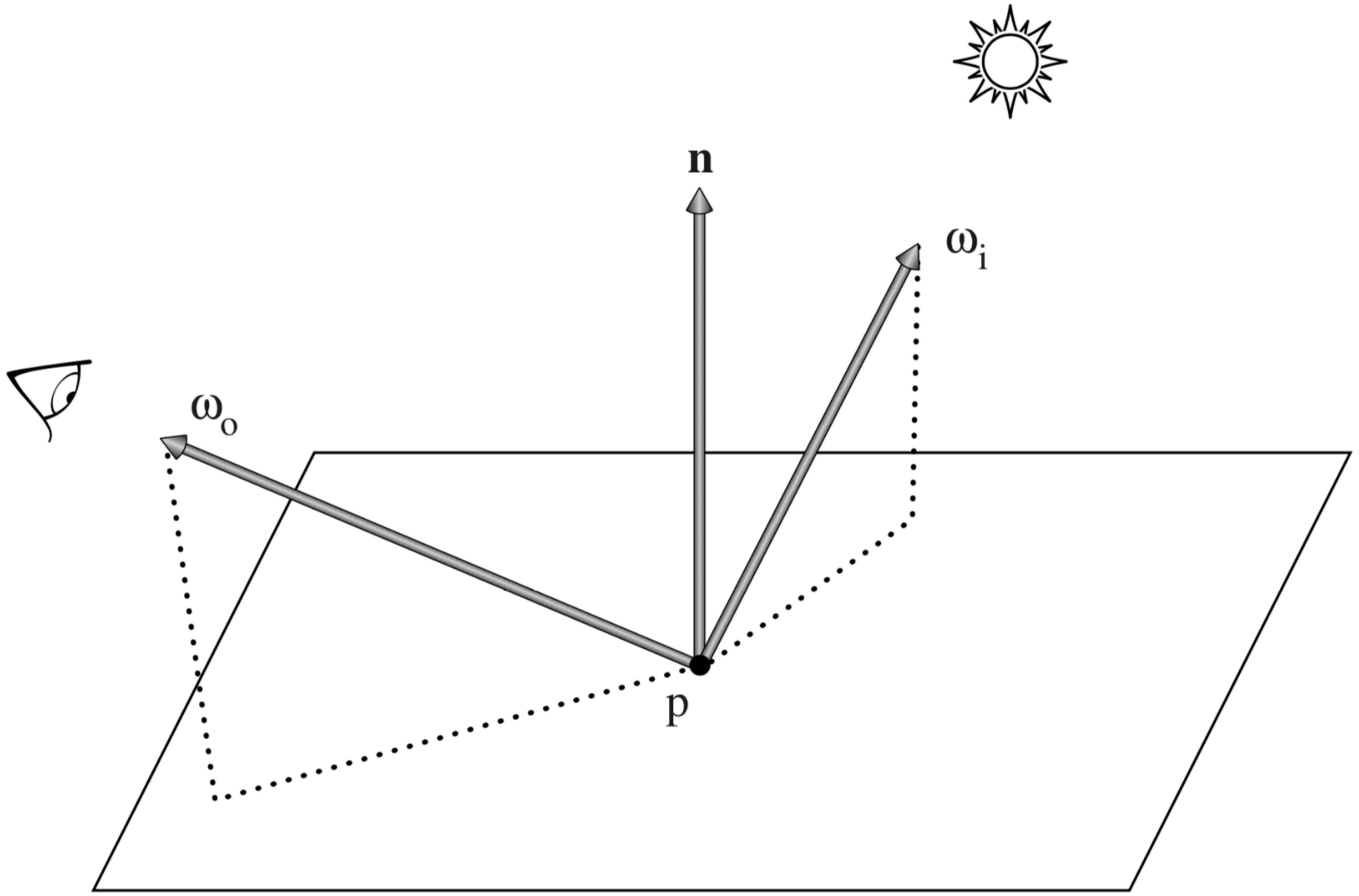
Due to the Strong law of large numbers, the arithmetic mean will converge to the expected value with probability 1 given enough samples:

$$\text{prob}\left\{ \lim_{N\to\infty} \mathbf{I}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)} = \mathbf{E}\left[\frac{f(x)}{p(x)}\right] = \int_Q f(x)dx \right\} = 1$$

UNIVERSITÄT DES SAARLANDES

Rendering Equation

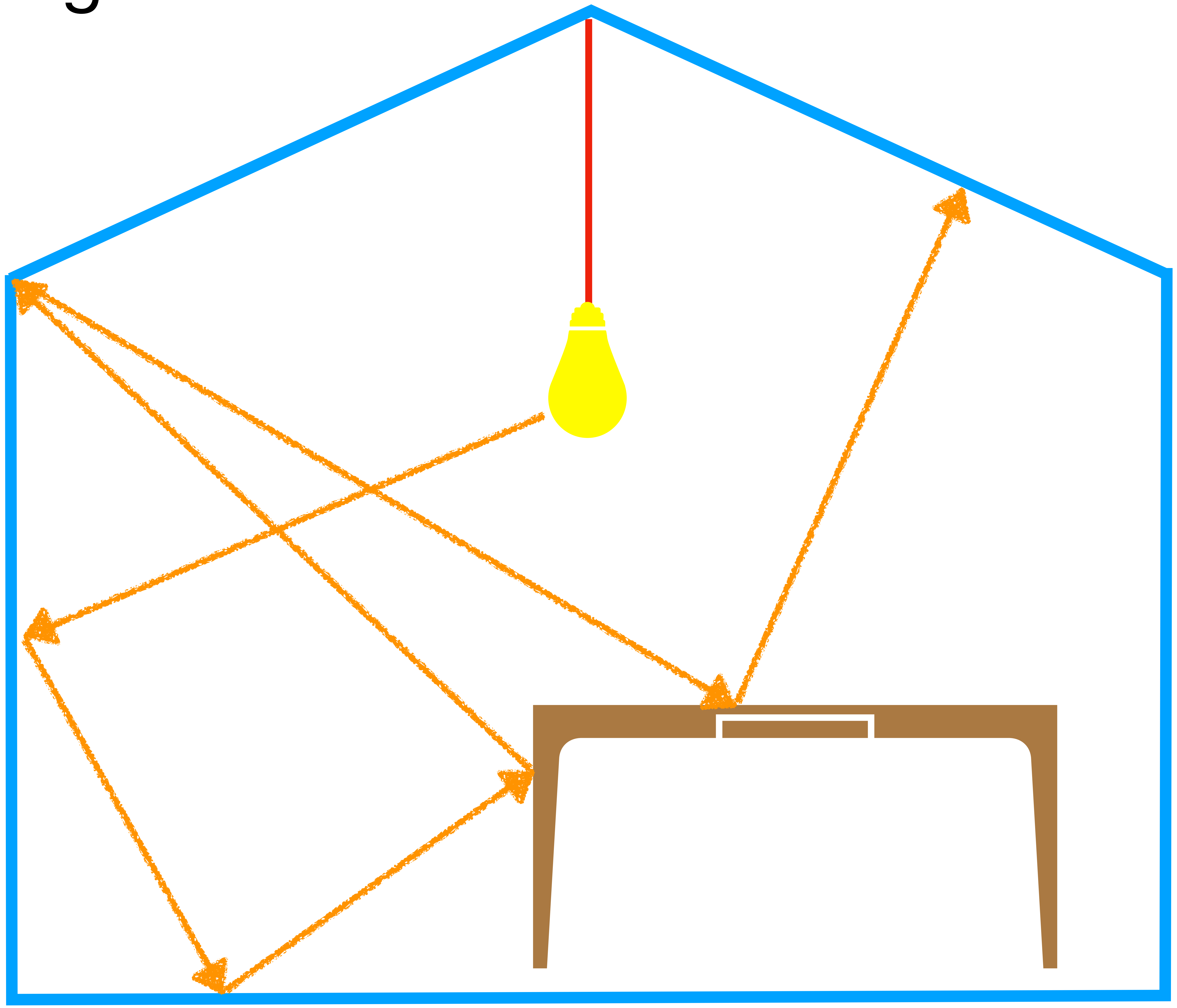
Scattering equation:



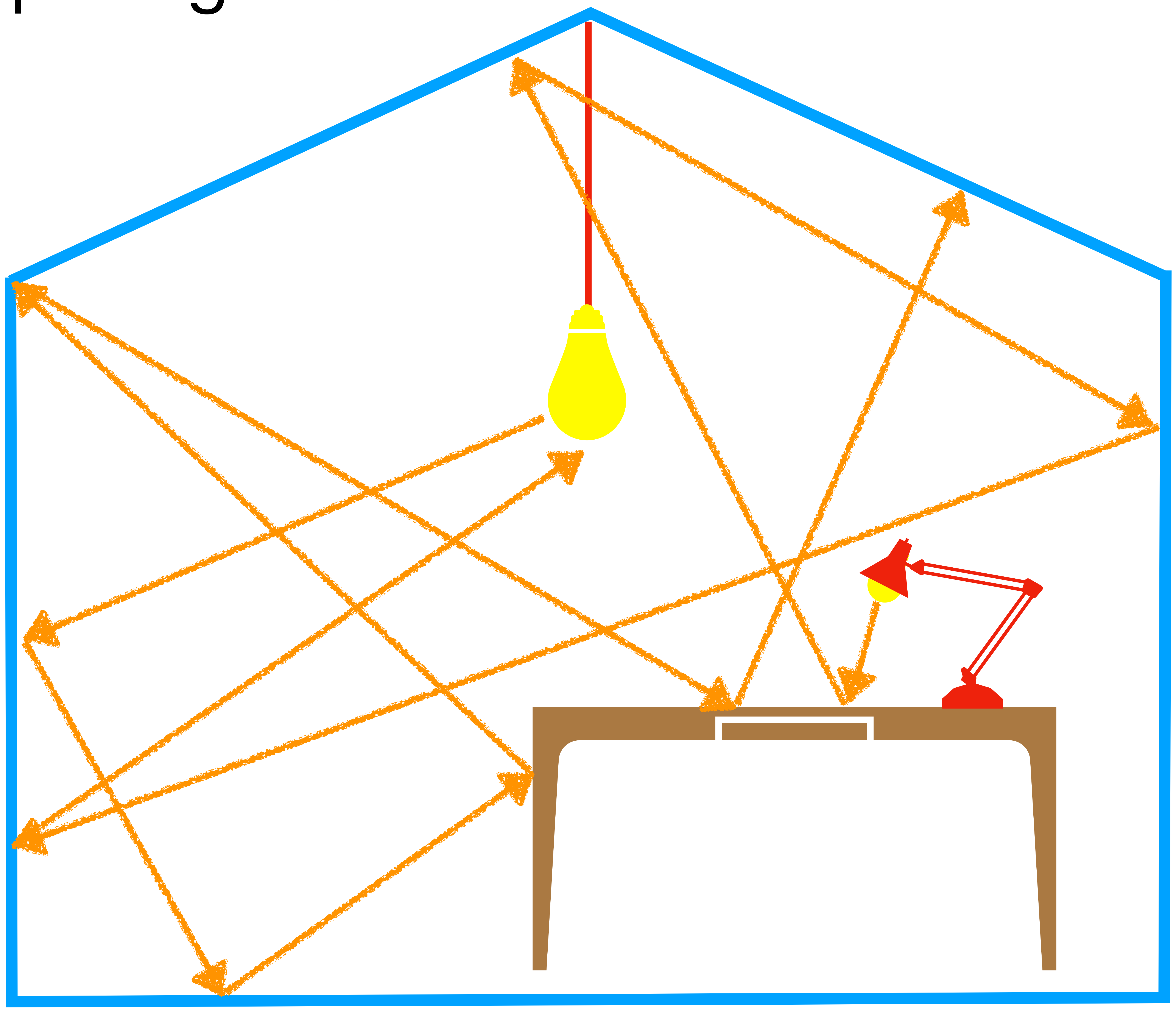
$$L_o(p, \omega_o) = \int_{S^2} f(p, \omega_o, \omega_i)\, L_i(p, \omega_i)\, |\cos\theta_i|\, d\omega_i$$

Image from PBRT 2016

Global Illumination: One Light Source



Global Illumination: Multiple Light Source



Error in Monte Carlo Estimation

$$\text{Error} = \text{Bias}^2 + \text{Variance}$$

Error in Monte Carlo Estimation

$$\text{Error} = \cancel{\text{Bias}^2} + \text{Variance}$$

- Monte Carlo estimation is unbiased due to it's "purely" stochastic nature
- We are left with variance, which is visible as stochastic unstructured noise in the rendered images

Error in Monte Carlo Estimation

$$\text{Error} = \text{Bias}^2 + \text{Variance}$$

- For biased techniques, it is important to have a consistent solution
    - This implies, the bias goes to zero with increase in sample count
    - Examples: Progressive photon mapping

Unbiased: Monte Carlo Estimator

$$\text{Error} = \mathbf{I}_N - \mathbf{I}$$

$$\text{Error} = \mathbf{I}_N - \int_Q f(x)dx$$

Unbiased: Monte Carlo Estimator

$$\text{Error} = \mathbf{I}_N - \int_Q f(x)dx$$

Bias by definition is the expected error:

$$\text{Bias} = \mathbf{E}[\text{Error}] = \mathbf{E}\left[\mathbf{I}_N - \int_Q f(x)dx\right]$$

$$\text{Bias} = \mathbf{E}\left[\mathbf{I}_N\right] - \left[\int_Q f(x)dx\right]$$

$$\text{Bias} = \mathbf{E}\left[\mathbf{I}_N\right] - \int_Q f(x)dx$$

Unbiased: Monte Carlo Estimator

$$\text{Bias} = \mathbf{E}\Big[\mathbf{I}_N\Big] - \int_Q f(x)dx$$

$$\mathbf{E}\Big[\mathbf{I}_N\Big] = \mathbf{E}\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}\Big] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\Big[\frac{f(x_i)}{p(x_i)}\Big] = \frac{1}{N}\sum_{i=1}^{N}\int_Q \frac{f(x)}{p(x)}p(x)dx$$

$$= \frac{1}{N}\sum_{i=1}^{N}\int_Q f(x)dx$$

$$= \int_Q f(x)dx$$

Unbiased: Monte Carlo Estimator

$$\text{Bias} = \mathbf{E}\left[\mathbf{I}_N\right] - \int_Q f(x)dx$$

$$\mathbf{E}\left[\mathbf{I}_N\right] = \int_Q f(x)dx$$

$$\text{Bias} = \mathbf{0}$$

Variance: Monte Carlo Estimator

For the variance of secondary Monte Carlo Estimator, the following holds:

$$\mathrm{Var}(\mathbf{I_N}) = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}(\mathbf{I}_1^i)$$

Variance: Monte Carlo Estimator

$$\mathrm{Var}(\mathbf{I_N}) = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}(\mathbf{I}_1^i)$$

$$\mathrm{Var}(\mathbf{I_N}) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}\right)$$

$$\boxed{\mathrm{Var}(aX) = a^2\mathrm{Var}(X)}$$

Variance: Monte Carlo Estimator

$$\mathrm{Var}(\mathbf{I_N}) = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}(\mathbf{I}_1^i)$$

$$\mathrm{Var}(\mathbf{I_N}) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}\right) = \frac{1}{N^2}\mathrm{Var}\left(\sum_{I=1}^{N}\frac{f(x_i)}{p(x_i)}\right)$$

$$\boxed{\mathrm{Var}(aX) = a^2\mathrm{Var}(X)}$$

$$= \frac{1}{N^2}\sum_{I=1}^{N}\mathrm{Var}\left(\frac{f(x_i)}{p(x_i)}\right)$$

Independent samples

$$= \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}(\mathbf{I}_1^i)$$

Convergence rate: MC Estimator

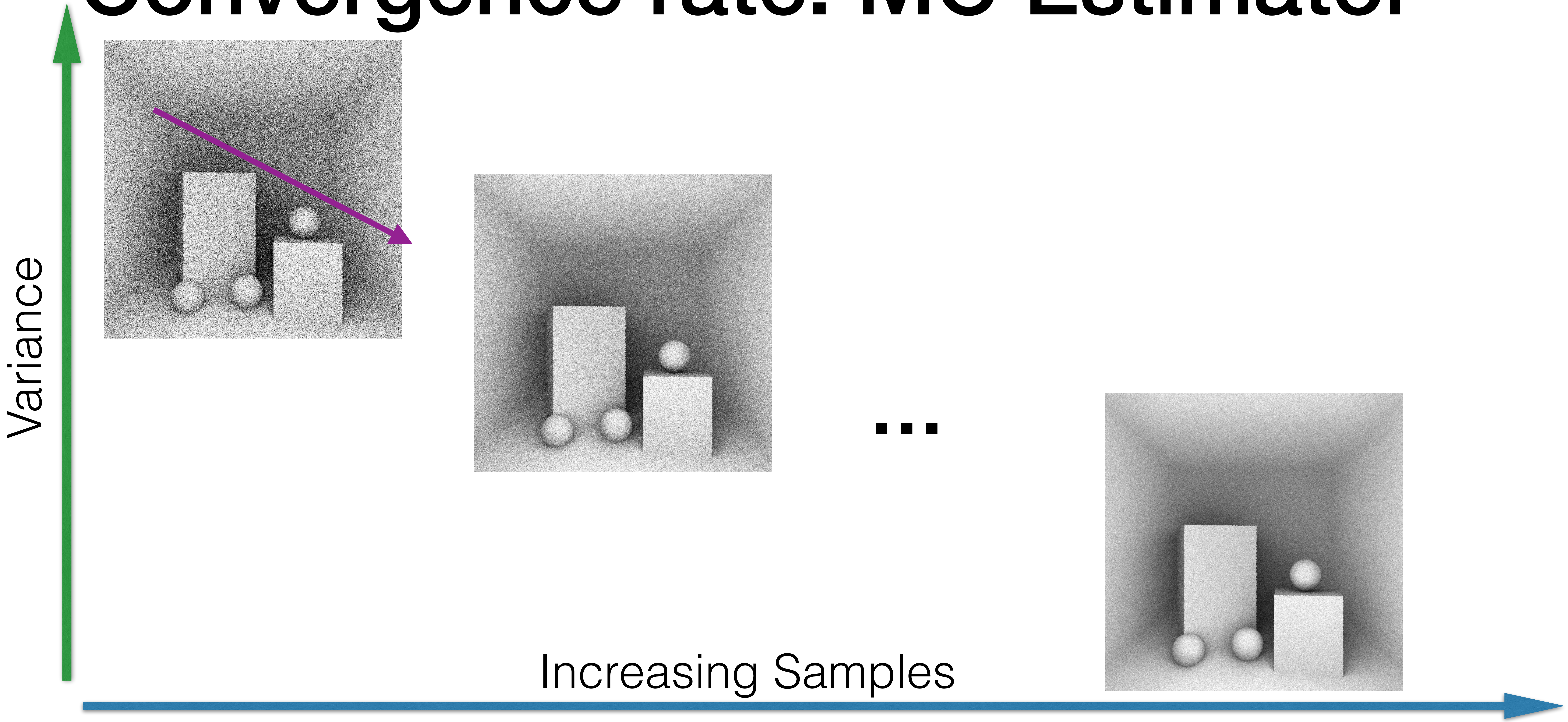
$$\mathrm{Var}(\mathbf{I_N}) = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}(\mathbf{I}_1^i)$$

$$\mathrm{Error} = \sigma(\mathbf{I}_N) = \frac{1}{\sqrt{N^2}}\sqrt{\mathrm{Var}(\mathbf{I}_1^i)}$$

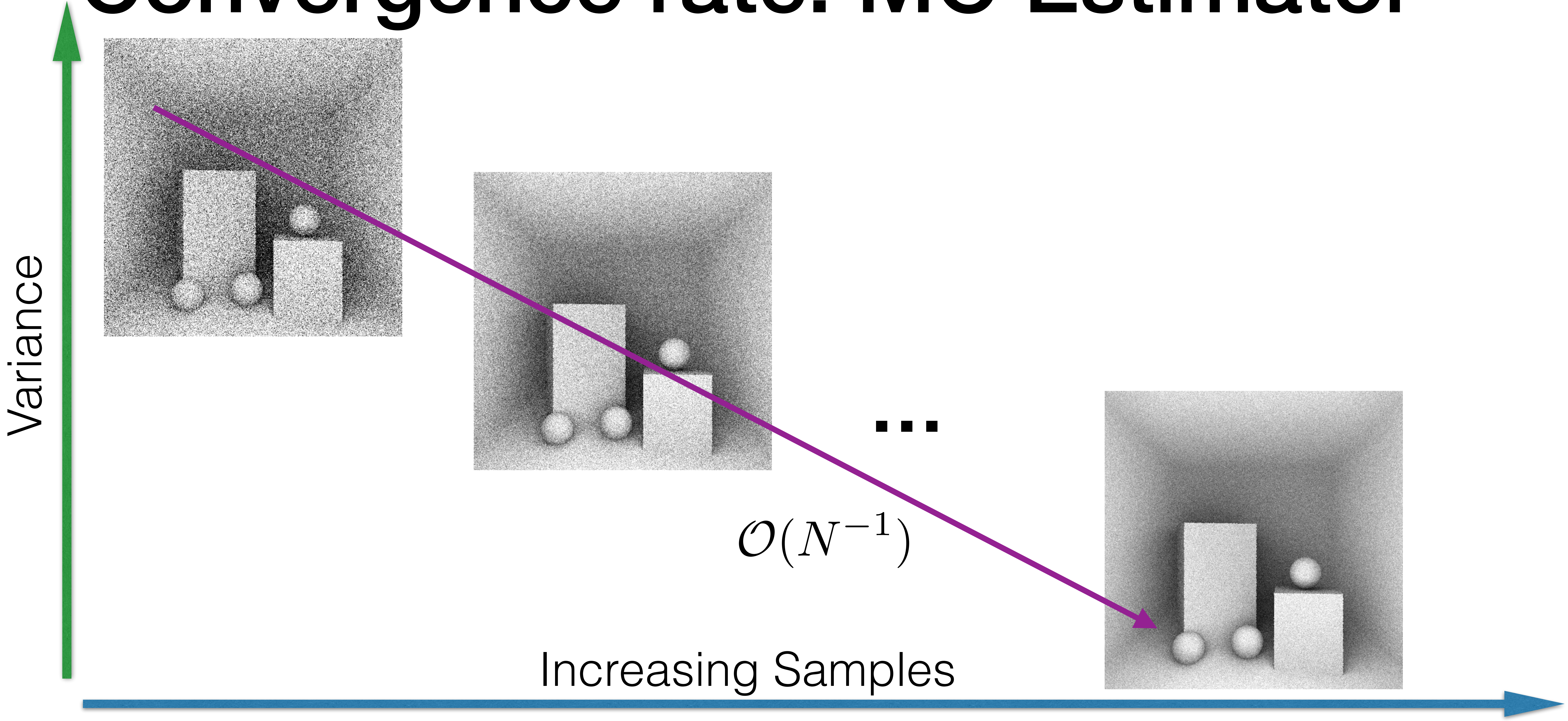
$$= \frac{1}{N}\sigma(\mathbf{I}_1^i)$$

$$\boxed{\sigma(X) = \sqrt{\mathrm{Var}(X)}}$$

Convergence rate: MC Estimator



Convergence rate: MC Estimator



Sampling Methods

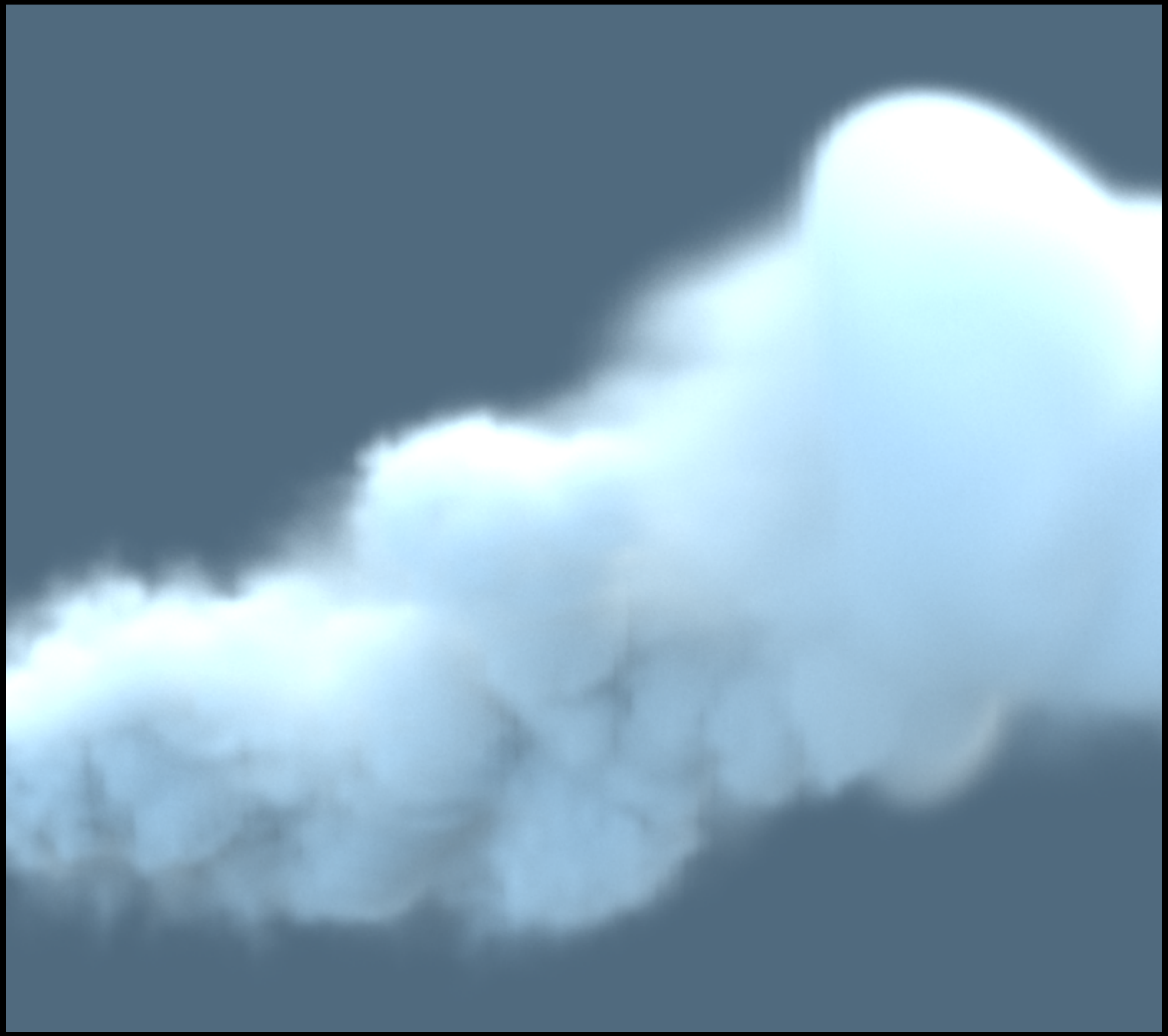
Sampling Methods

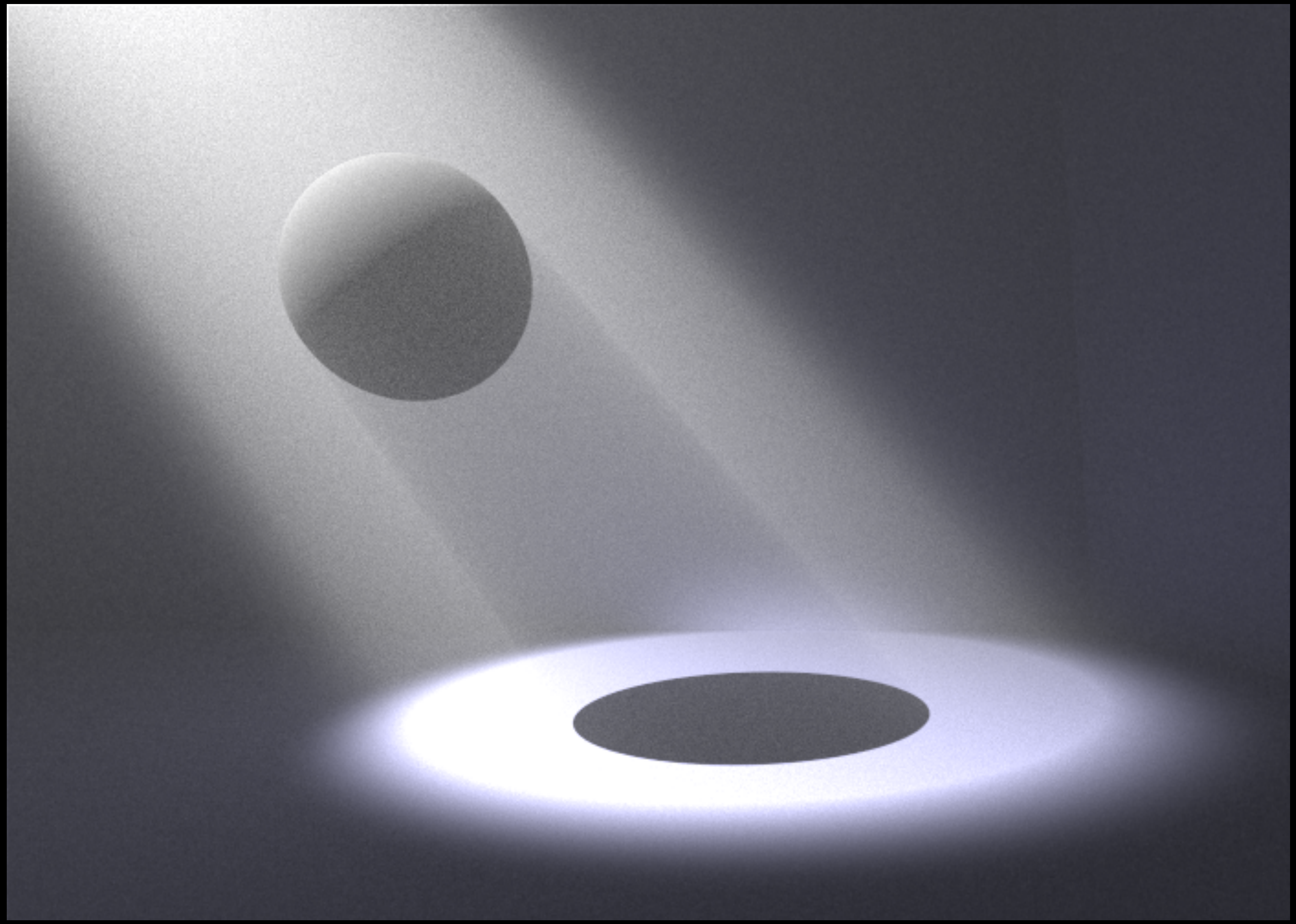
- Inversion methods
- Acceptance-rejection methods
- Metropolis sampling (later)
- Transforming distributions

Inversion Method

- Compute the CDF $P(x) = \int_0^x p(z)dz$
- Compute the inverse CDF $P^{-1}(x)$
- Obtain a uniformly distributed random number $\xi \in [0,1)$
- Compute $X_i = P^{-1}(\xi)$

Rendering participating media





Inversion Method

$$p(x) \propto e^{-ax}$$

$$p(x) = ce^{-ax}$$

$$\boxed{\int_0^\infty ce^{-ax}dx = \frac{c}{a} = 1}$$

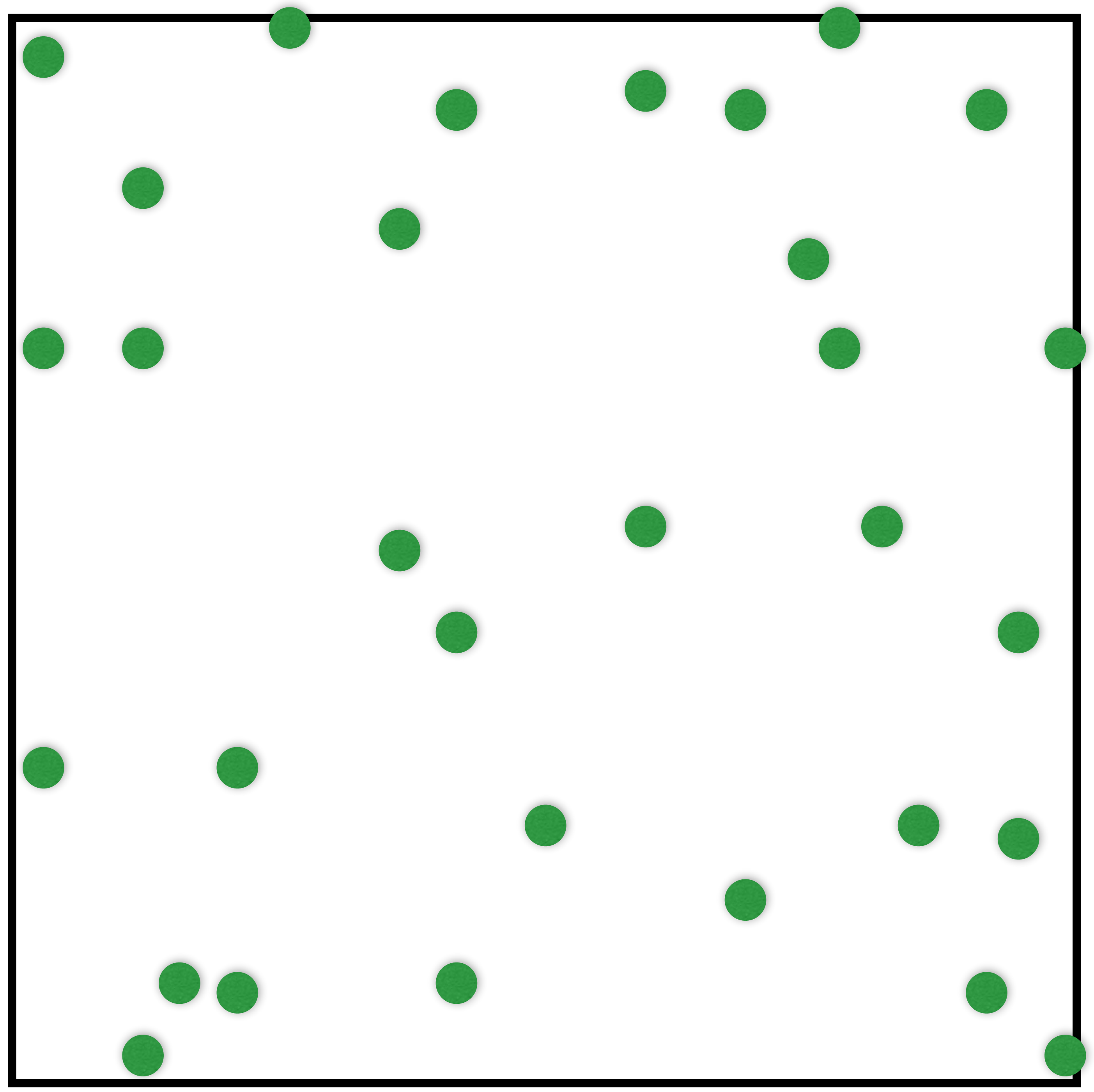
$$P(x) = \int_0^x ce^{-ax}dx = 1 - e^{-ax} = \xi$$

$$P^{-1}(x) = \frac{ln(1-\xi)}{a}$$

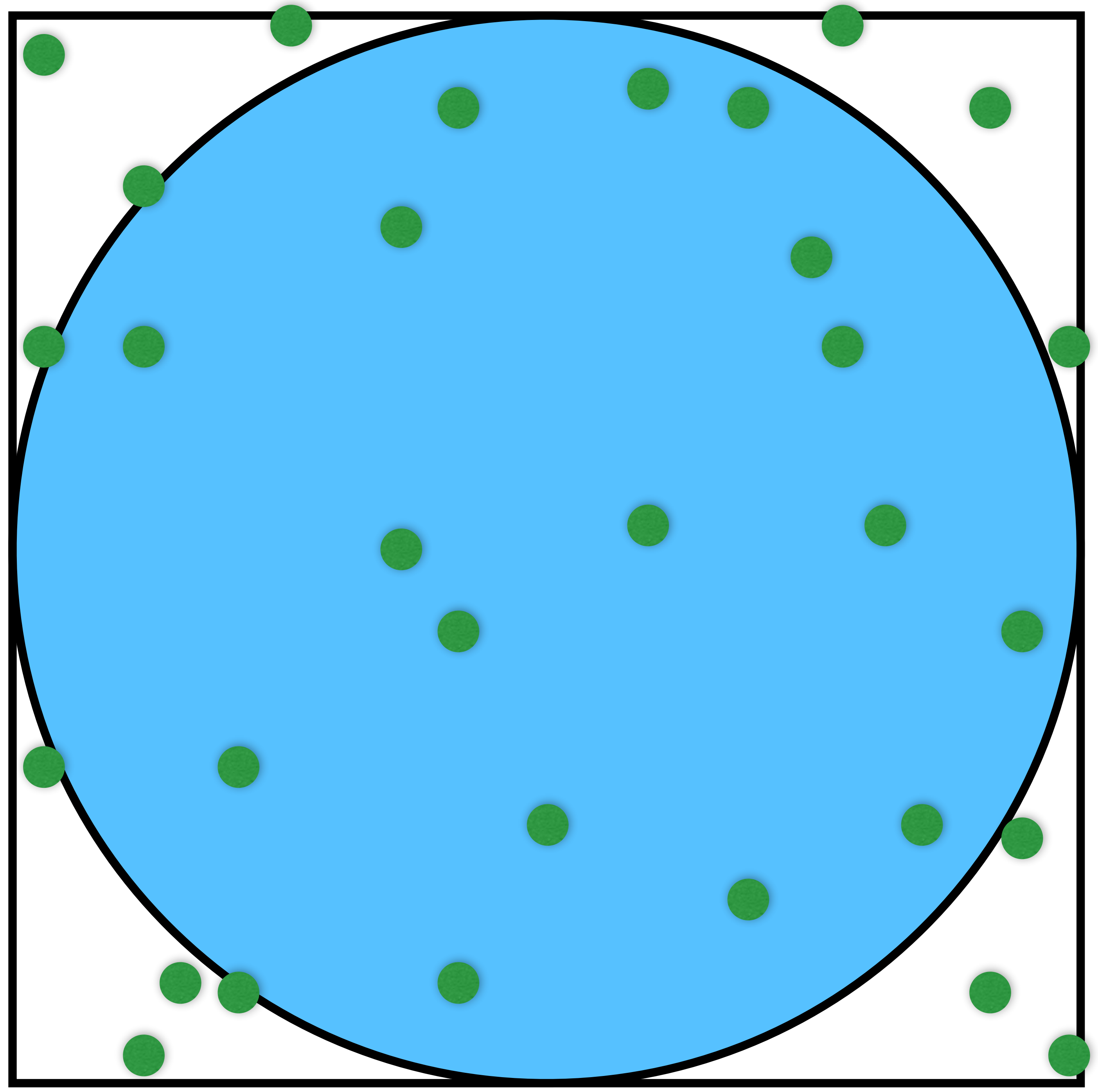
$$\boxed{P^{-1}(x) = \frac{ln(1-x)}{a}}$$

$$P^{-1}(x) = \frac{ln(\xi)}{a}$$

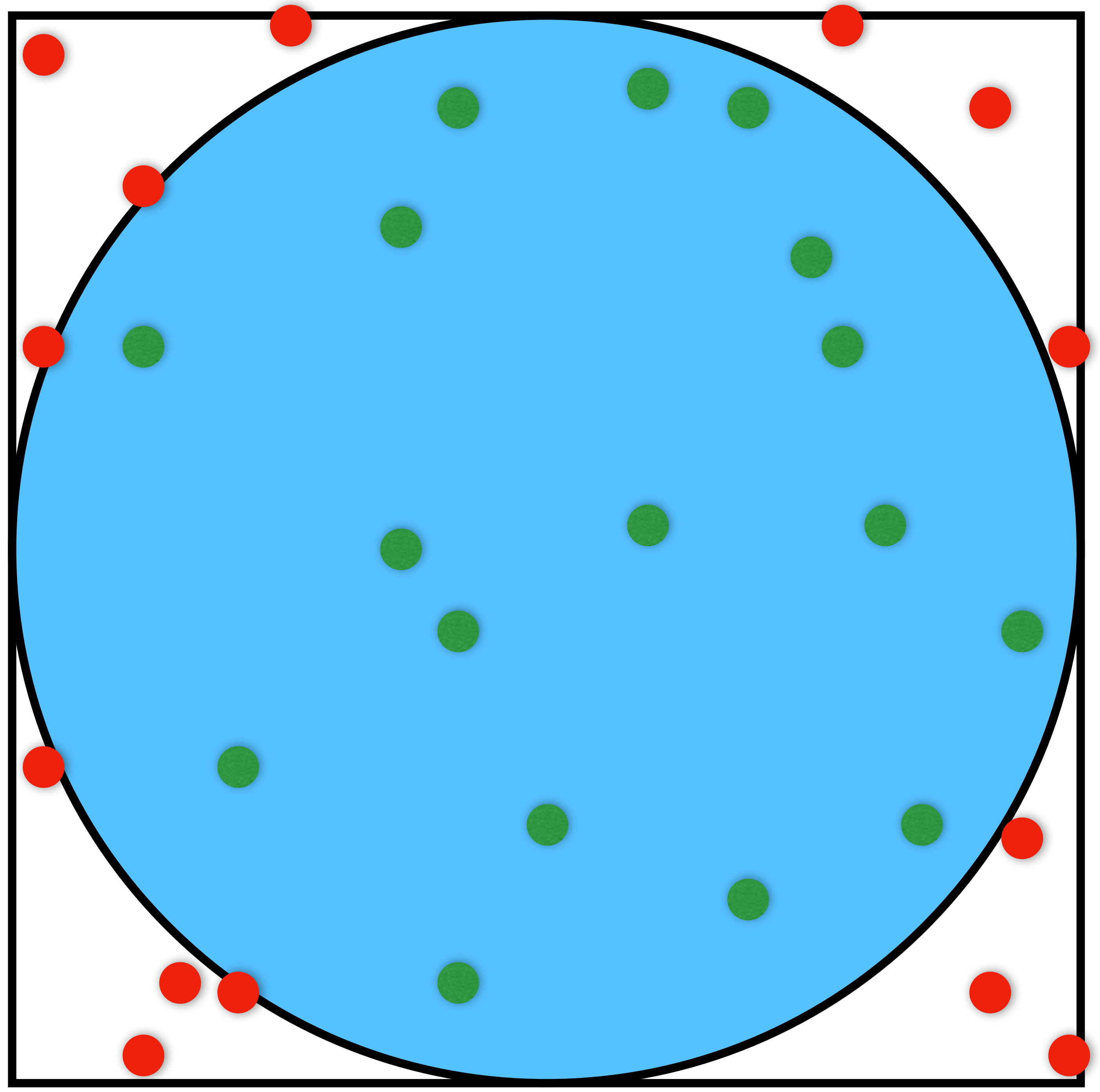
Rejection Sampling Method



Rejection Sampling Method



Rejection Sampling Method



- Many samples are wasted
- Very costly
- Not possible for arbitrary domains

Transformation Method

- General question: which distributions results when we transform samples from an arbitrary distributions to some other distribution with a function f .

$$X_i \sim p_x(x)$$

$$Y_i = y(X_i)$$

What is the distribution of Y_i ?

Transformation Method

- The function $y(x)$ must be a one-to-one transformation
  - It's derivative must either be strictly > 0 or strictly < 0

$$\mathrm{prob}\{Y \leq y(x)\} = \mathrm{prob}\{X \leq x\}$$

Transformation Method

$$\text{prob}\{Y \leq y(x)\} = \text{prob}\{X \leq x\}$$

$$P_y(y) = P_y(y(x)) = P_x(x)$$

This relationship between CDFs directly leads to the relationship between their PDFs:

$$p_y(y)\frac{dy}{dx} = p_x(x)$$

$$p_y(y) = \left(\frac{dy}{dx}\right)^{-1} p_x(x)$$

Transformation Method

$$p_y(y) = \left(\frac{dy}{dx}\right)^{-1} p_x(x)$$

In general, the derivative is strictly positive or negative, and the relationship between the densities is:

$$p_y(y) = \left|\frac{dy}{dx}\right|^{-1} p_x(x)$$

Transformation Method

$$p_y(y) = \left|\frac{dy}{dx}\right|^{-1} p_x(x)$$

How can we use this formula ?

$$p_x(x) = 2x \qquad x \in [0,1]$$

$$Y = \sin X$$

$$\frac{dy}{dx} = \cos x$$

$$p_y(y) = \frac{p_x(x)}{|\cos x|} = \frac{2x}{\cos x} = \frac{2\arcsin y}{\sqrt{1-y^2}}$$

Transformation Method

- Usually we have some PDF that we want to sample from, not a given transformation

- For example, we might have given: $X \sim p_x(x)$ and we would like to compute $Y \sim p_y(y)$

$$P_y(y) = P_x(x) \qquad\qquad y(x) = P_y^{-1}(P_x(x))$$

- This is a generalization of the inversion method.

Transformation in Multiple dimensions

- Suppose we have an s-dimensional X with density function p_X
- Now let $Y = T(X)$ where T is a bijection.

$$p_y(y) = p_y(T(x)) = \frac{p_x(x)}{|J_T(x)|}$$

$$J_T(x) = \begin{pmatrix} \partial T_1/\partial x_1 & \cdots & \partial T_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial T_n/\partial x_1 & \cdots & \partial T_n/\partial x_n \end{pmatrix}$$

Polar Coordinates

$$x = r\cos\theta$$
$$y = r\sin\theta$$

Suppose we draw samples from some density $p(r,\theta)$

What is the corresponding density $p(x,y)$?

$$J_T = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

$$p(x,y) = p(r,\theta)/J_T$$

$$p(x,y) = p(r,\theta)/r$$

Spherical Coordinates

$$x = r \sin\theta \cos\phi$$
$$y = r \sin\theta \sin\phi$$
$$z = r \cos\theta,$$

$$|J_T| = r^2 \sin\theta$$

$$p(r, \theta, \phi) = r^2 \sin\theta \; p(x, y, z)$$

Spherical Coordinates

Spherical coordinates

$$x = r\,\sin\theta\,\cos\phi$$
$$y = r\,\sin\theta\,\sin\phi$$
$$z = r\,\cos\theta,$$

$$|J_T| = r^2\,\sin\theta$$

$$d\omega = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$$

$$p(r,\theta,\phi) = r^2\sin\theta\;p(x,y,z)$$

$$Pr\,\{\omega\in\Omega\} = \int_\Omega p(\omega)\,\mathrm{d}\omega$$

$$p(\theta,\phi)\,\mathrm{d}\theta\,\mathrm{d}\phi = p(\omega)\,\mathrm{d}\omega$$
$$p(\theta,\phi) = \sin\theta\;p(\omega)$$

Uniformly sampling a hemisphere

Here, the task is to choose a direction on the hemisphere uniformly w.r.t. solid angle.
Using the fact that, PDF must integrate to one over its domain:

$$\int_{\mathcal{H}^2} p(\omega)\,\mathrm{d}\omega = 1 \Rightarrow c \int_{\mathcal{H}^2} \mathrm{d}\omega = 1 \Rightarrow c = \frac{1}{2\pi}$$

$$p(\omega) = 1/(2\pi)$$

$$p(\theta, \phi) = \sin\theta/(2\pi)$$

Marginal density function: $p(\theta) = \int_0^{2\pi} p(\theta, \phi)\,\mathrm{d}\phi = \int_0^{2\pi} \frac{\sin\theta}{2\pi}\,\mathrm{d}\phi = \sin\theta$

Conditional density function: $p(\phi|\theta) = \frac{p(\theta, \phi)}{p(\theta)} = \frac{1}{2\pi}$

Uniformly sampling a hemisphere

Corresponding CDFs:

$$P(\theta) = \int_0^\theta \sin\theta' \, d\theta' = 1 - \cos\theta$$

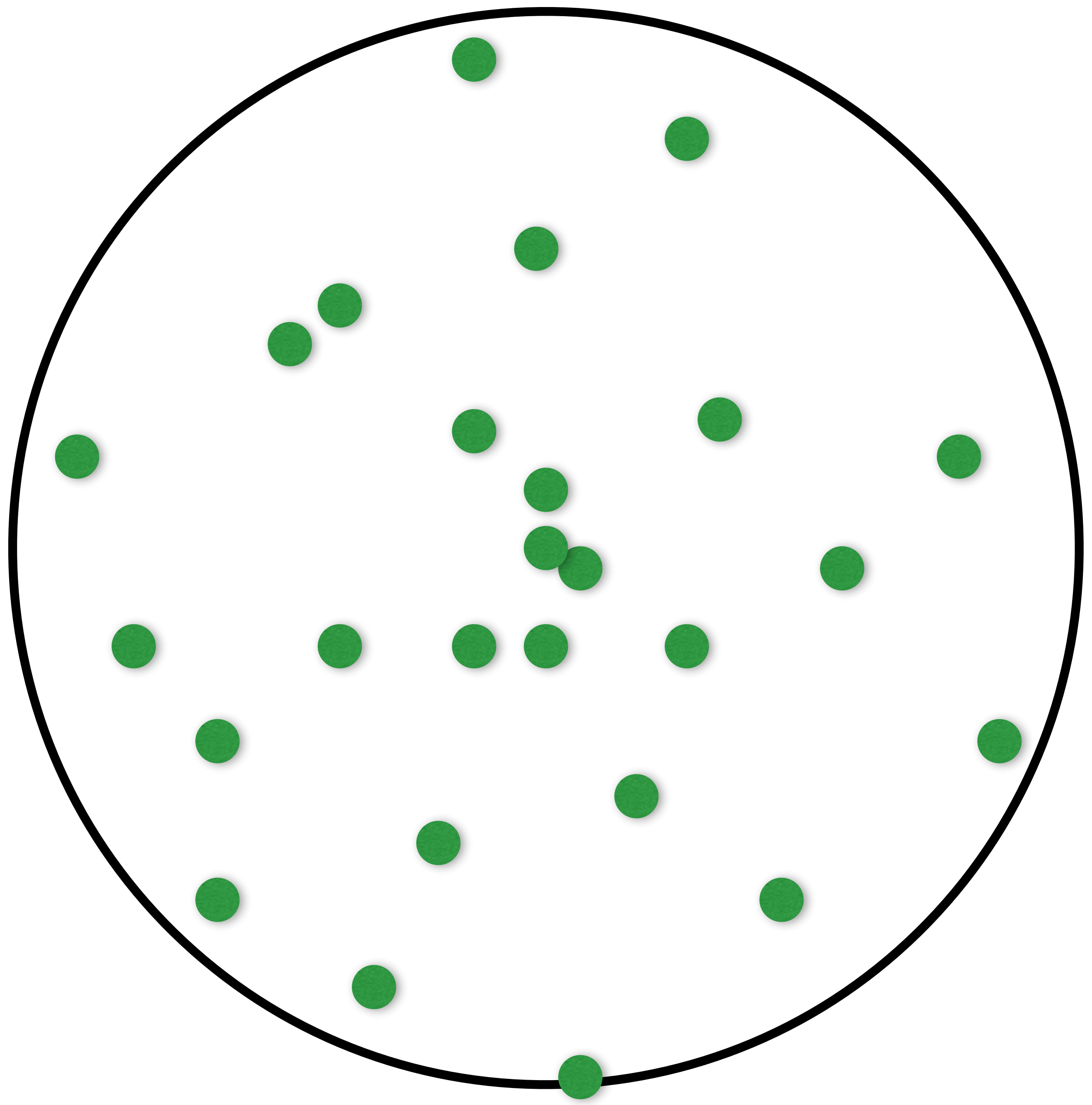
$$P(\phi|\theta) = \int_0^\phi \frac{1}{2\pi} \, d\phi' = \frac{\phi}{2\pi}.$$

Inverting these functions is straightforward, and here we can safely write:

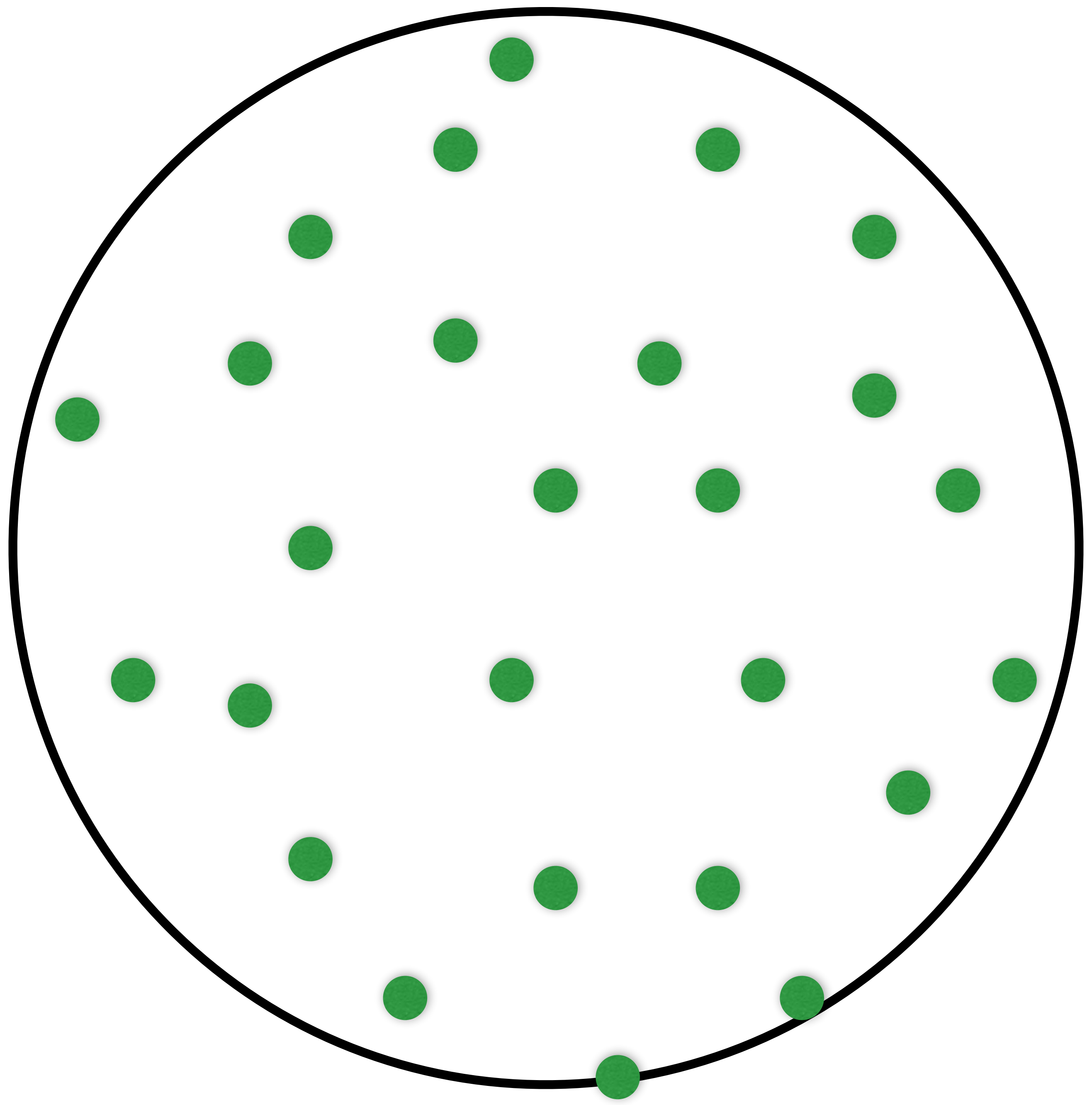
$$\theta = \cos^{-1}\xi_1$$
$$\phi = 2\pi\xi_2.$$

$$x = \sin\theta\cos\phi = \cos(2\pi\xi_2)\sqrt{1-\xi_1^2}$$
$$y = \sin\theta\sin\phi = \sin(2\pi\xi_2)\sqrt{1-\xi_1^2}$$
$$z = \cos\theta = \xi_1.$$

Uniformly sampling a disk



$$r = \xi_1, \theta = 2\pi \xi_2$$



Correct PDF ???

Uniformly sampling a disk

$$p(x, y) = 1/\pi$$

$$p(r, \theta) = r/\pi$$

Marginal density function: $$p(r) = \int_0^{2\pi} p(r, \theta)\,\mathrm{d}\theta = 2r$$

Conditional density function: $$p(\theta|r) = \frac{p(r, \theta)}{p(r)} = \frac{1}{2\pi}.$$

$$r = \sqrt{\xi_1}$$

$$\theta = 2\pi\xi_2.$$

Variance Reduction Techniques

Variance Reduction Techniques

- Importance Sampling
- Multiple Importance Sampling
- Control Variates
- Stratified Sampling

Variance reduction: Importance sampling

$$\mathbf{I}_N = \frac{1}{N} \frac{f(\vec{x}_i)}{p(\vec{x}_i)}$$

- Importance Sampling doesn't always reduce variance.
- The pdf $p(\vec{x})$ must be carefully chosen to gain improvements

Variance reduction: Importance sampling

$$\mathbf{I}_N = \frac{1}{N}\frac{f(\vec{x_i})}{p(\vec{x_i})}$$

$$p(\vec{x}) \propto f(\vec{x})$$

$$p(\vec{x}) = cf(\vec{x})$$

$$c = \frac{1}{\int_{-\infty}^{\infty} f(\vec{x})} d\vec{x}$$

$$\boxed{\int_{-\infty}^{\infty} p(\vec{x})d\vec{x} = 1}$$

$$\int_{-\infty}^{\infty} cf(\vec{x})d\vec{x} = 1$$

this seems like a no-op since the PDF computation requires the integral of the function that we are interested in estimating.

Variance reduction: Importance sampling

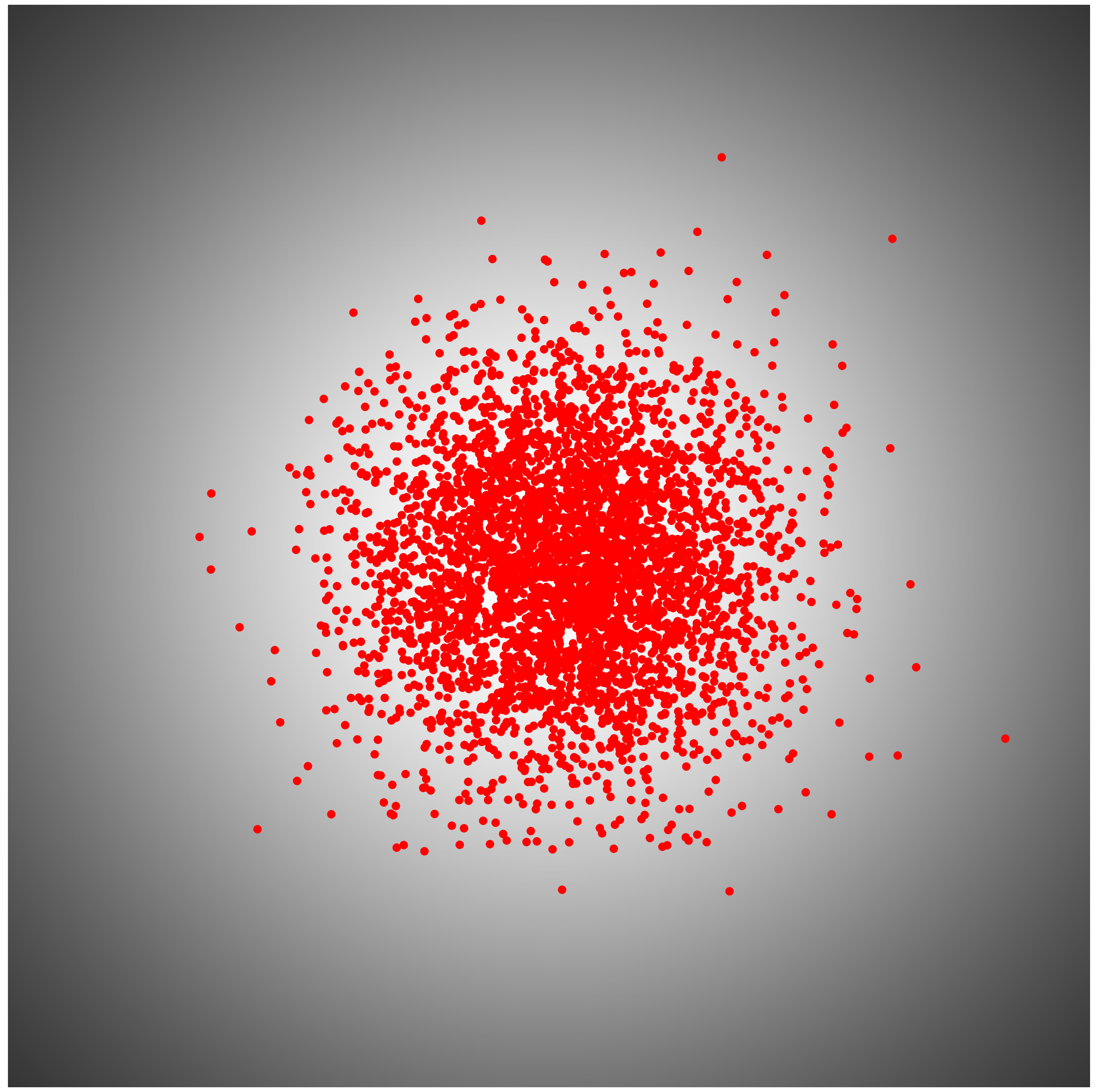
$$\mathbf{I}_N = \frac{1}{N}\frac{f(\vec{x_i})}{p(\vec{x_i})}$$

$$p(\vec{x}) = \frac{f(\vec{x})}{\int_{-\infty}^{\infty} f(\vec{x})} d\vec{x}$$

$$\mathbf{I}_N = \int_{-\infty}^{\infty} f(\vec{x}) d\vec{x}$$

- However, this is a very special case that we are encountering here.
- This is referred to as Perfect Importance Sampling, for which the variance is zero.

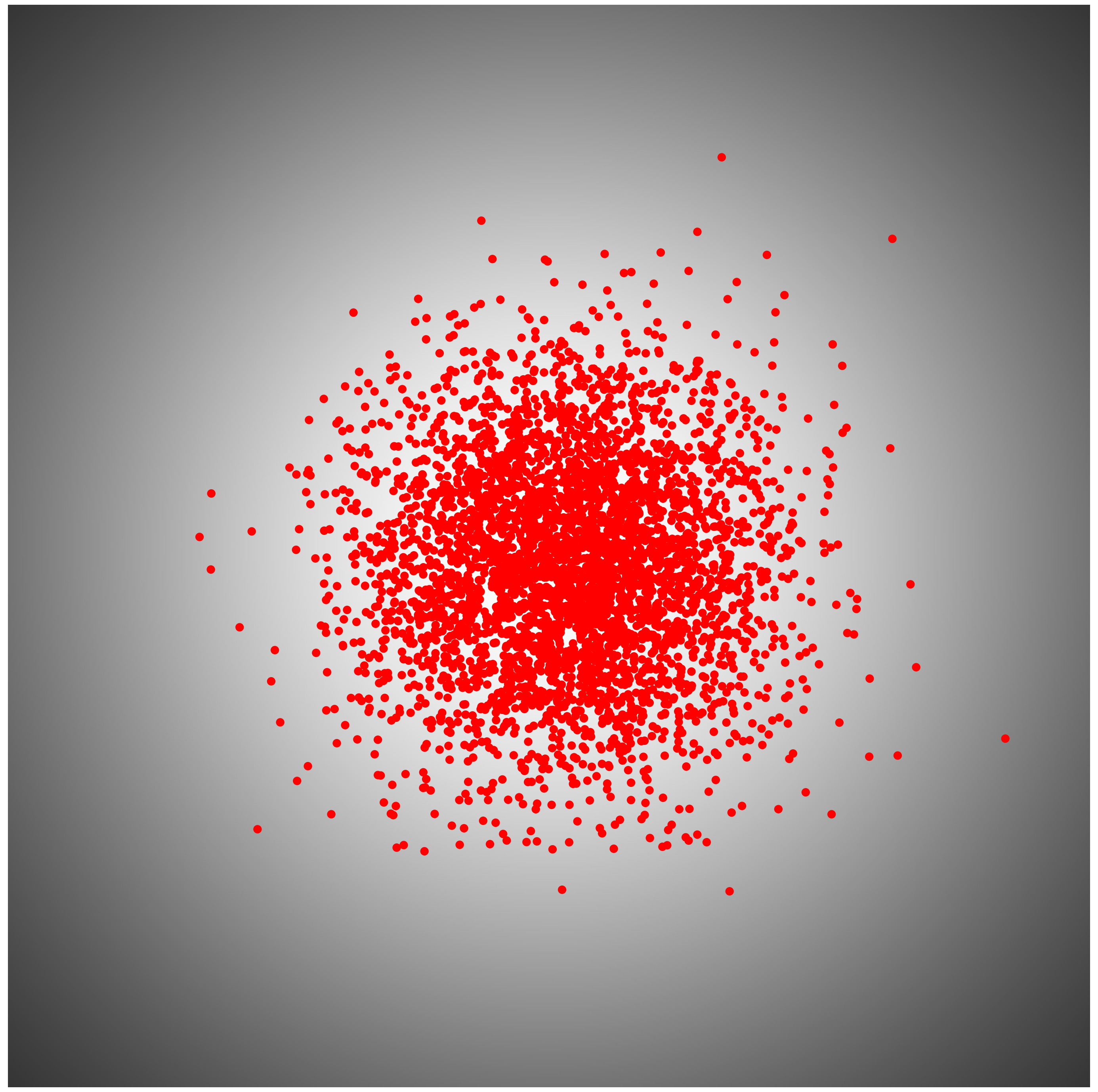
Variance reduction: Importance sampling



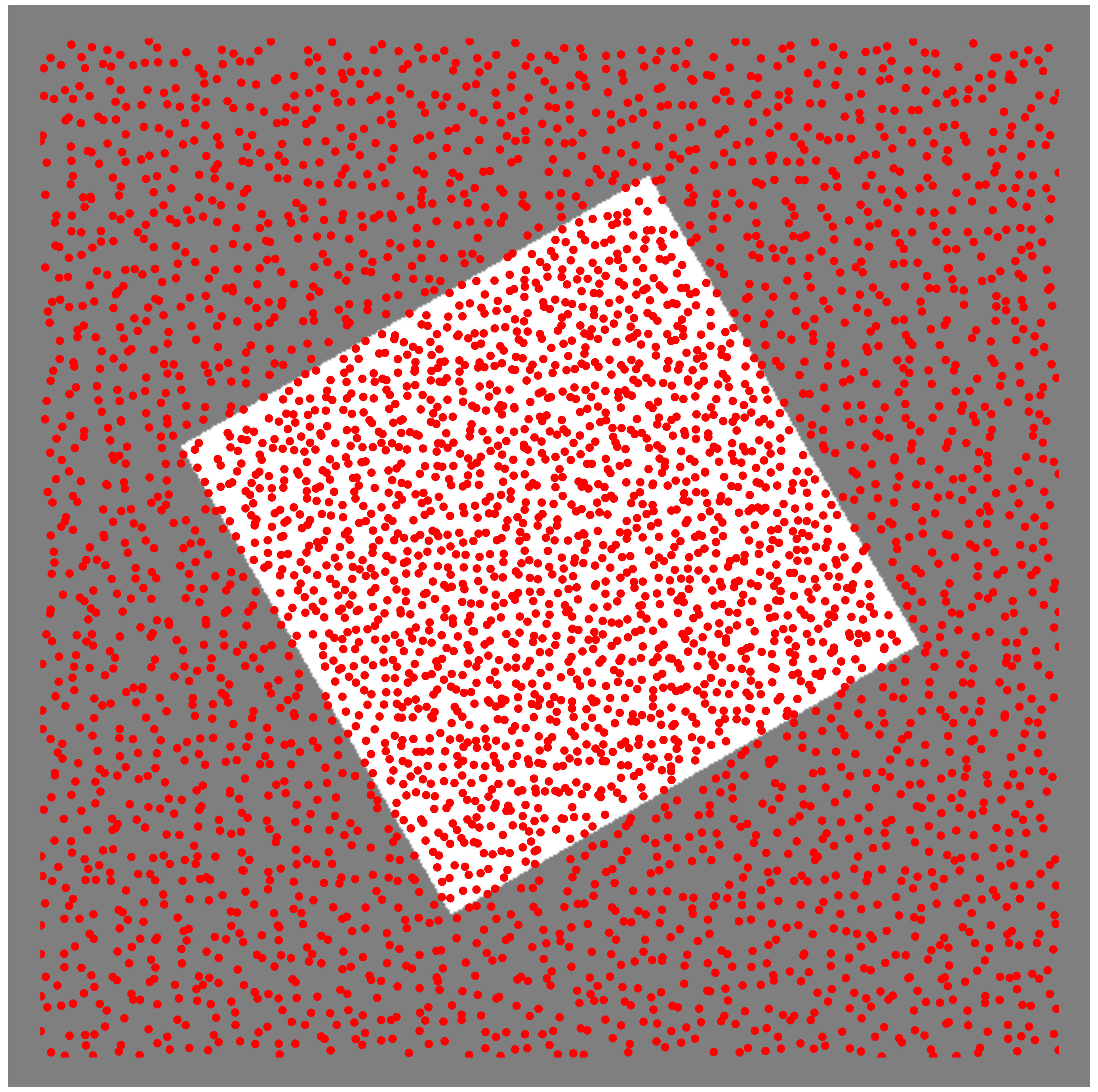
$f(\vec{x})$

Examples of **perfect importance sampling** for which the variance is zero

Variance reduction: Importance sampling



$f(\vec{x})$

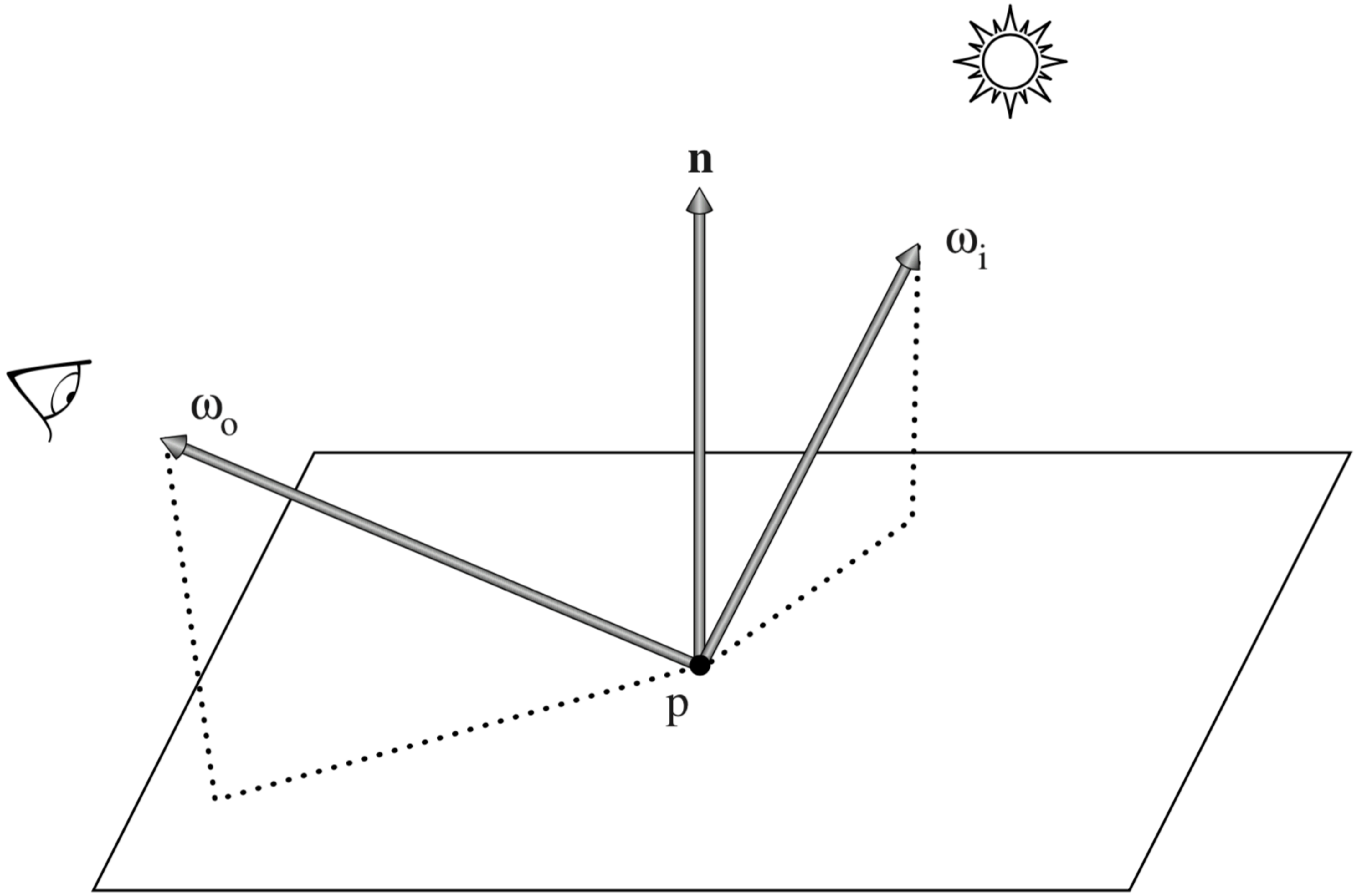


$g(\vec{x})$

Examples of **perfect importance sampling** for which the variance is zero

Variance reduction: Importance sampling

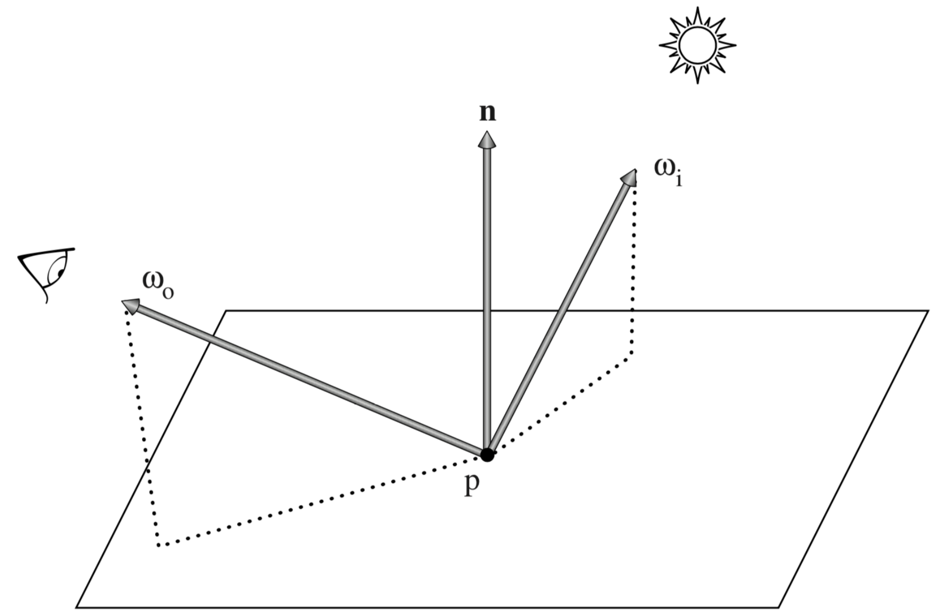
Scattering equation:



$$L_o(\mathrm{p}, \omega_o) = \int_{\mathcal{S}^2} f(\mathrm{p}, \omega_o, \omega_i)\, L_i(\mathrm{p}, \omega_i)\, |\cos\theta_i|\, \mathrm{d}\omega_i$$

Image from PBRT 2016

Variance reduction: Importance sampling



Scattering equation:

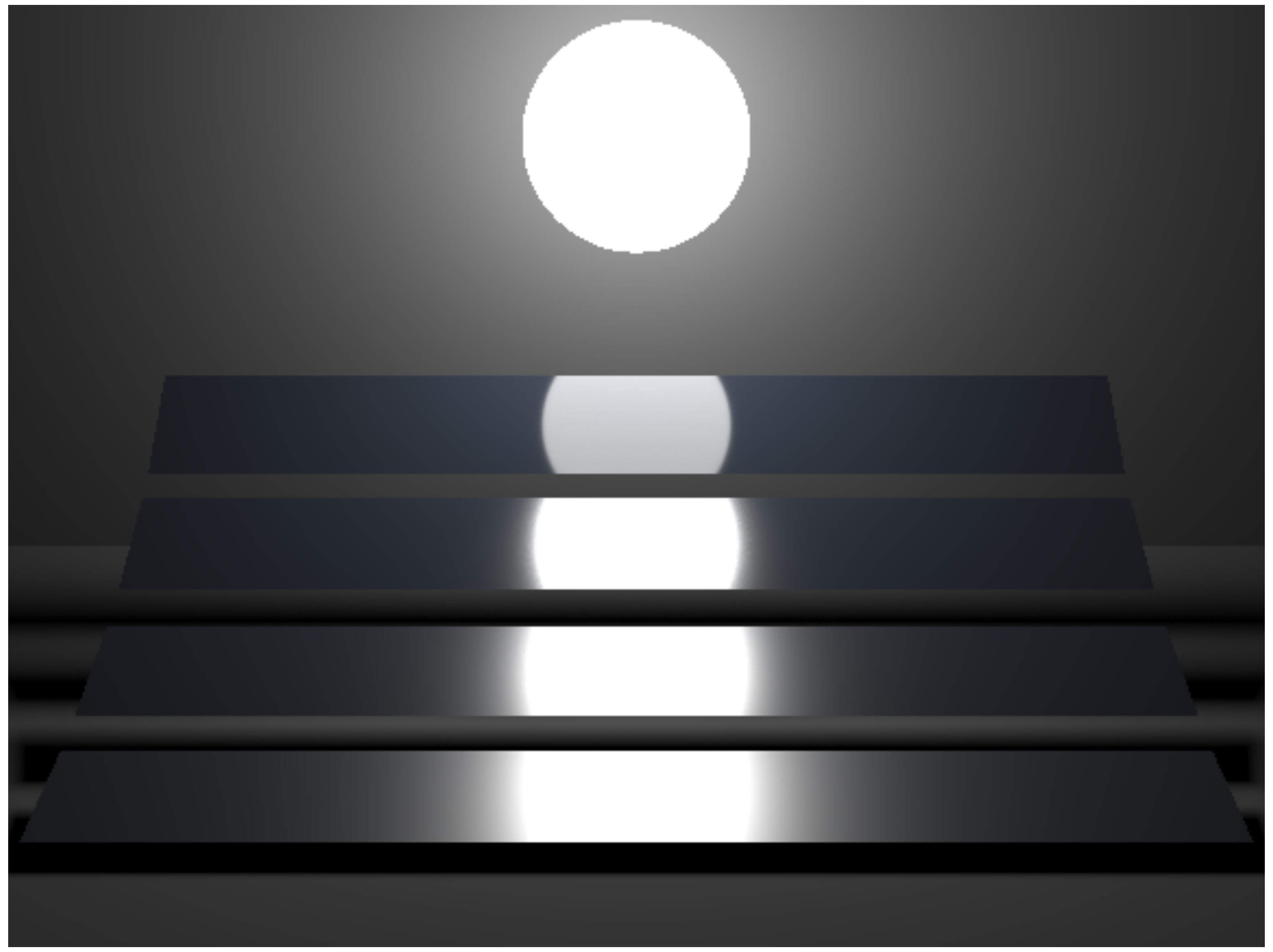
$$L_o(\mathrm{p}, \omega_o) = \int_{\mathbb{S}^2} f(\mathrm{p}, \omega_o, \omega_i)\, L_i(\mathrm{p}, \omega_i)\, |\cos\theta_i|\, \mathrm{d}\omega_i$$

$$\approx \frac{1}{N} \sum_{j=1}^{N} \frac{f(\mathrm{p}, \omega_o, \omega_j)\, L_i(\mathrm{p}, \omega_j)\, |\cos\theta_j|}{p(\omega_j)}$$

$$p(\omega) \propto \cos\theta$$

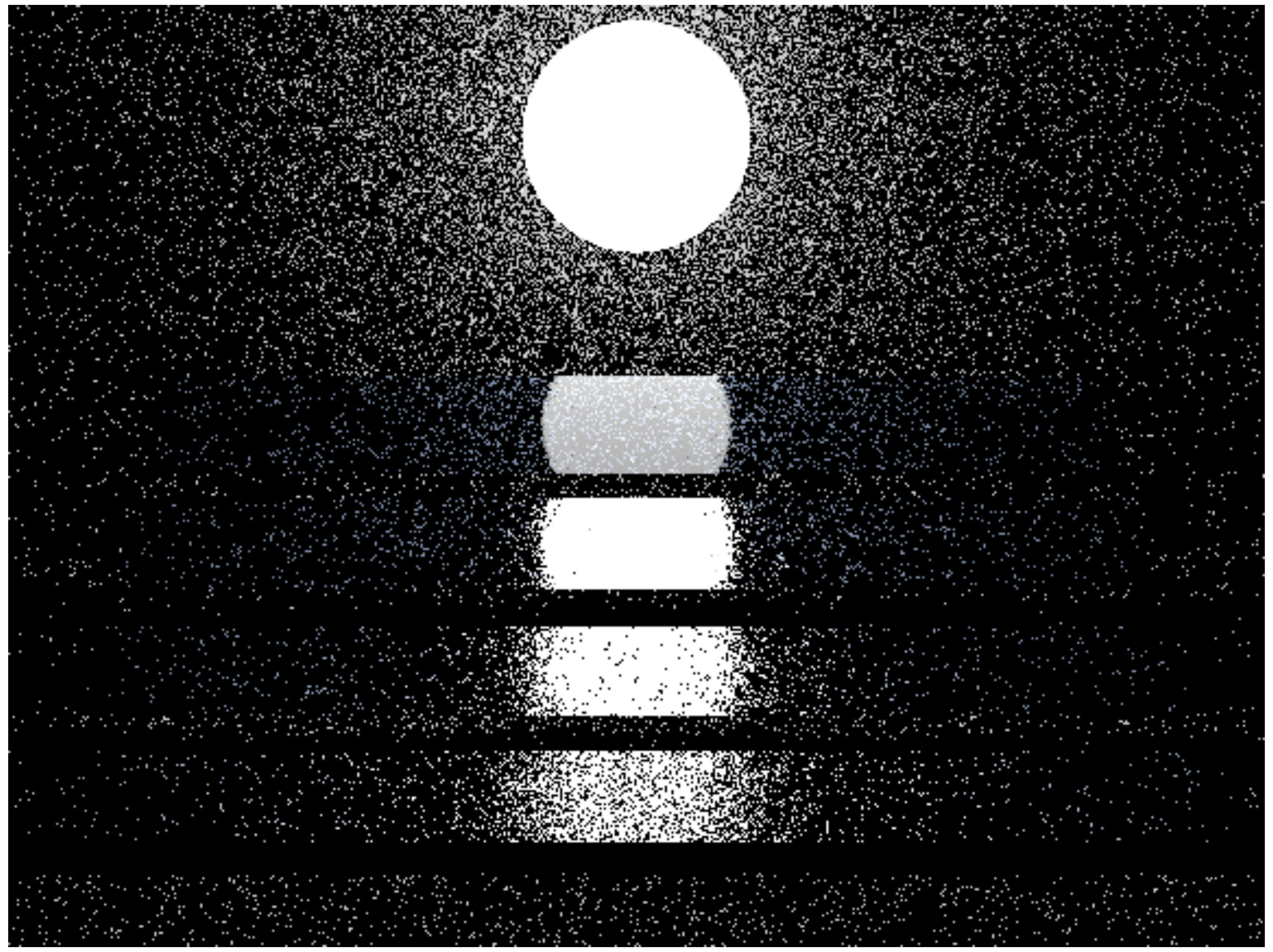
Cosine weighted spherical/hemispherical sampling

Variance reduction: Importance sampling



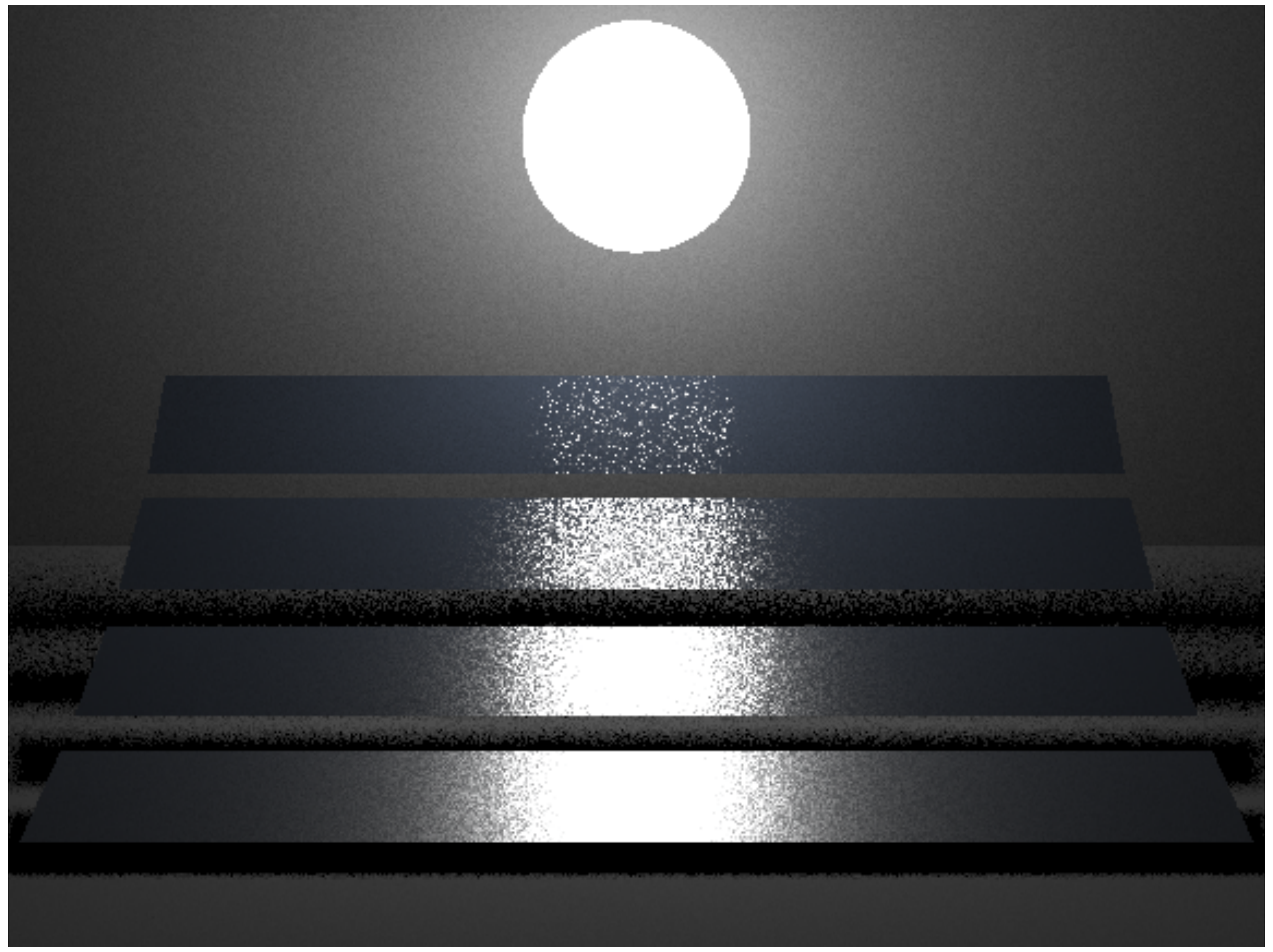
Reference image

N = 1024 spp



BSDF importance sampling

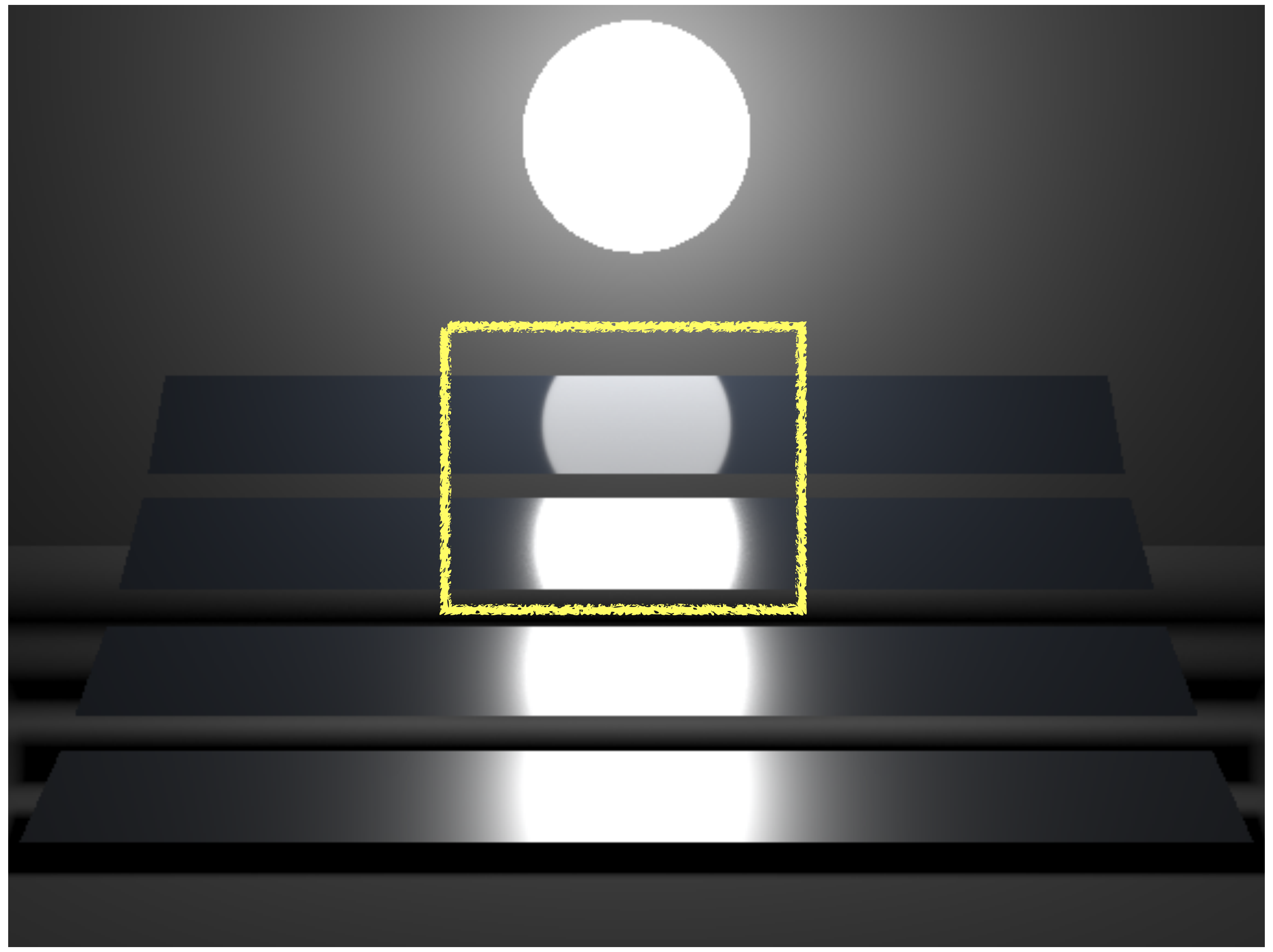
N = 4 spp



Light importance sampling

N = 4 spp

Variance reduction: Importance sampling

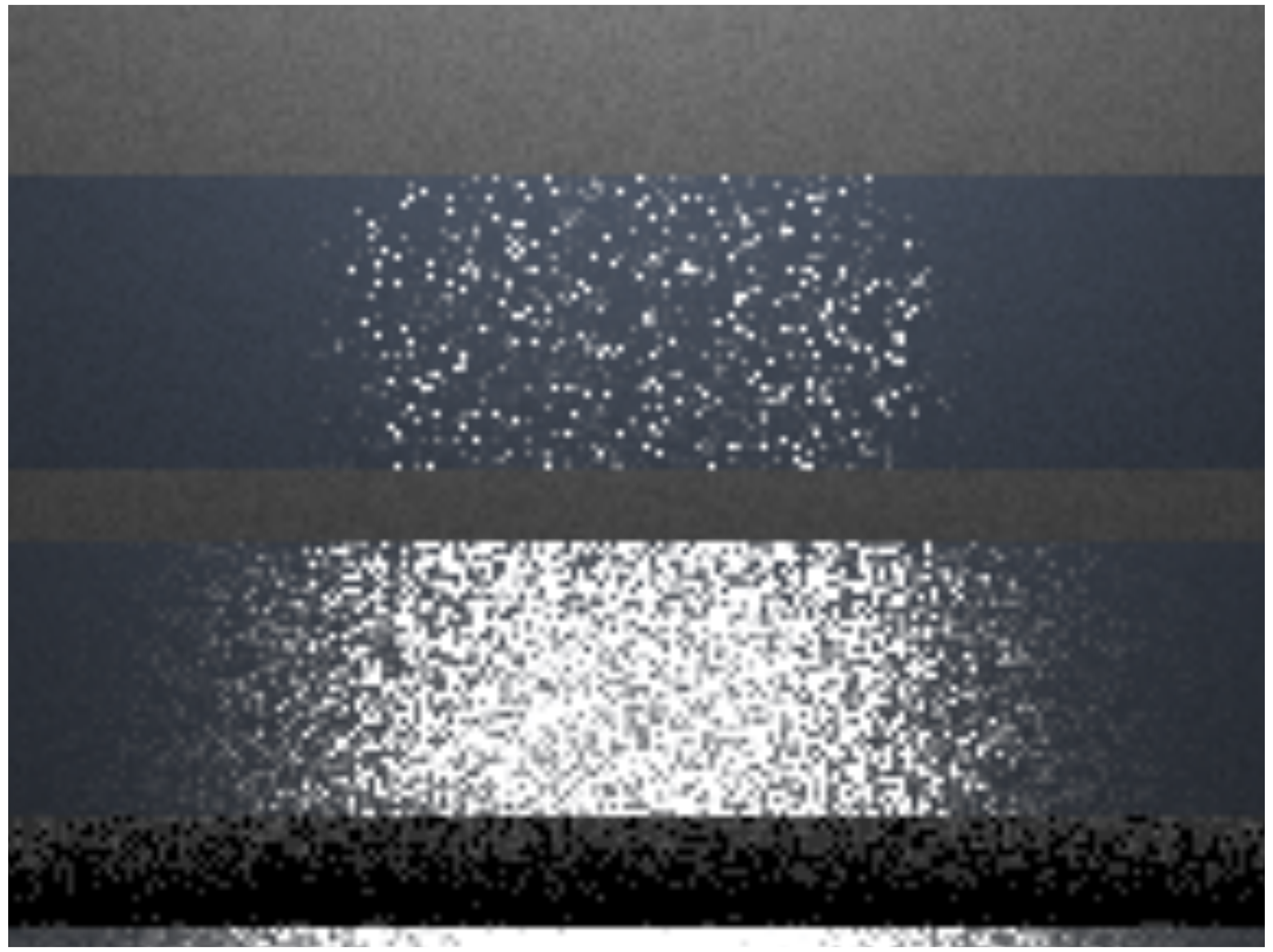


Reference image

N = 1024 spp

BSDF importance sampling

N = 4 spp

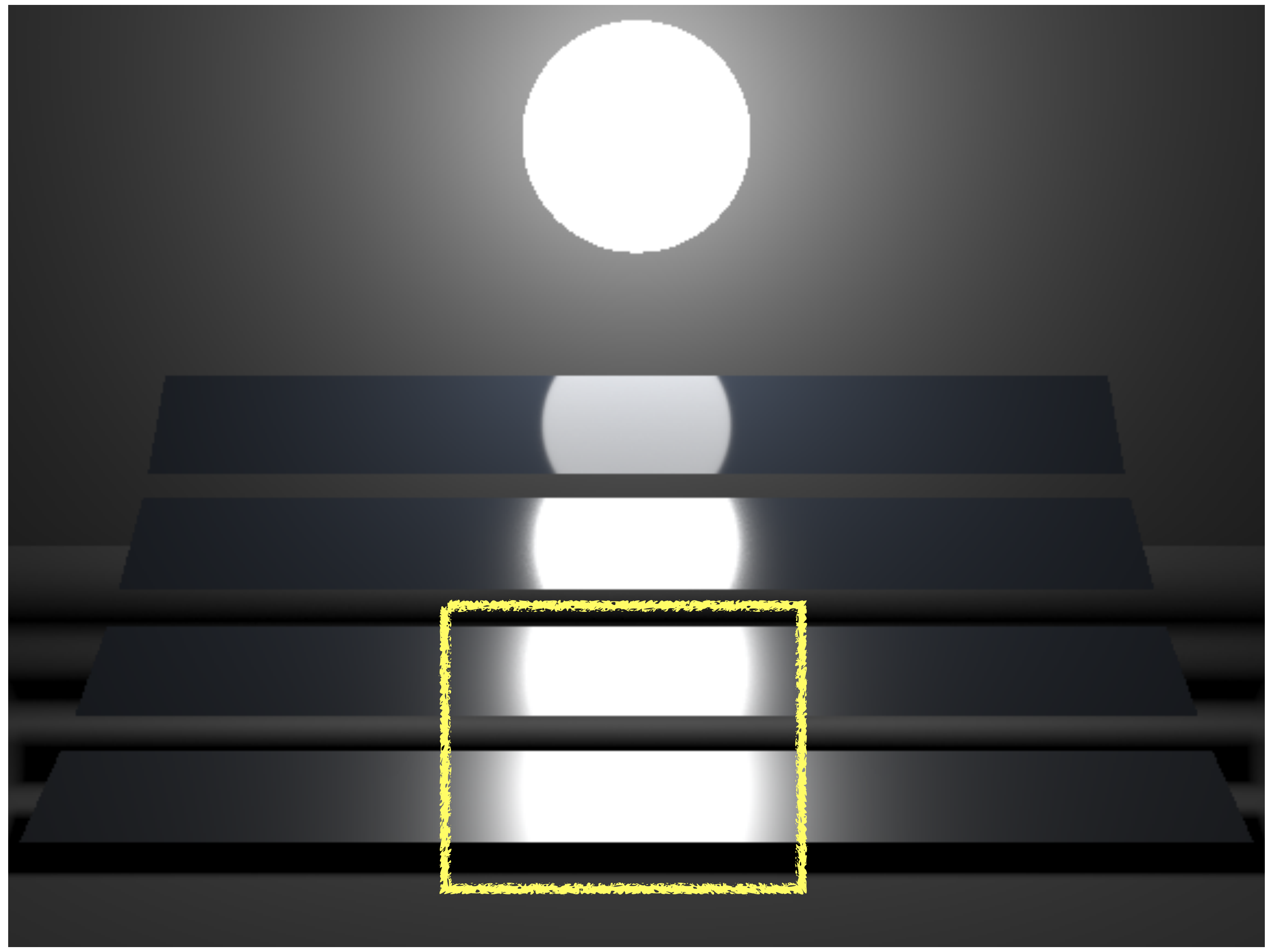


Light importance sampling

N = 4 spp

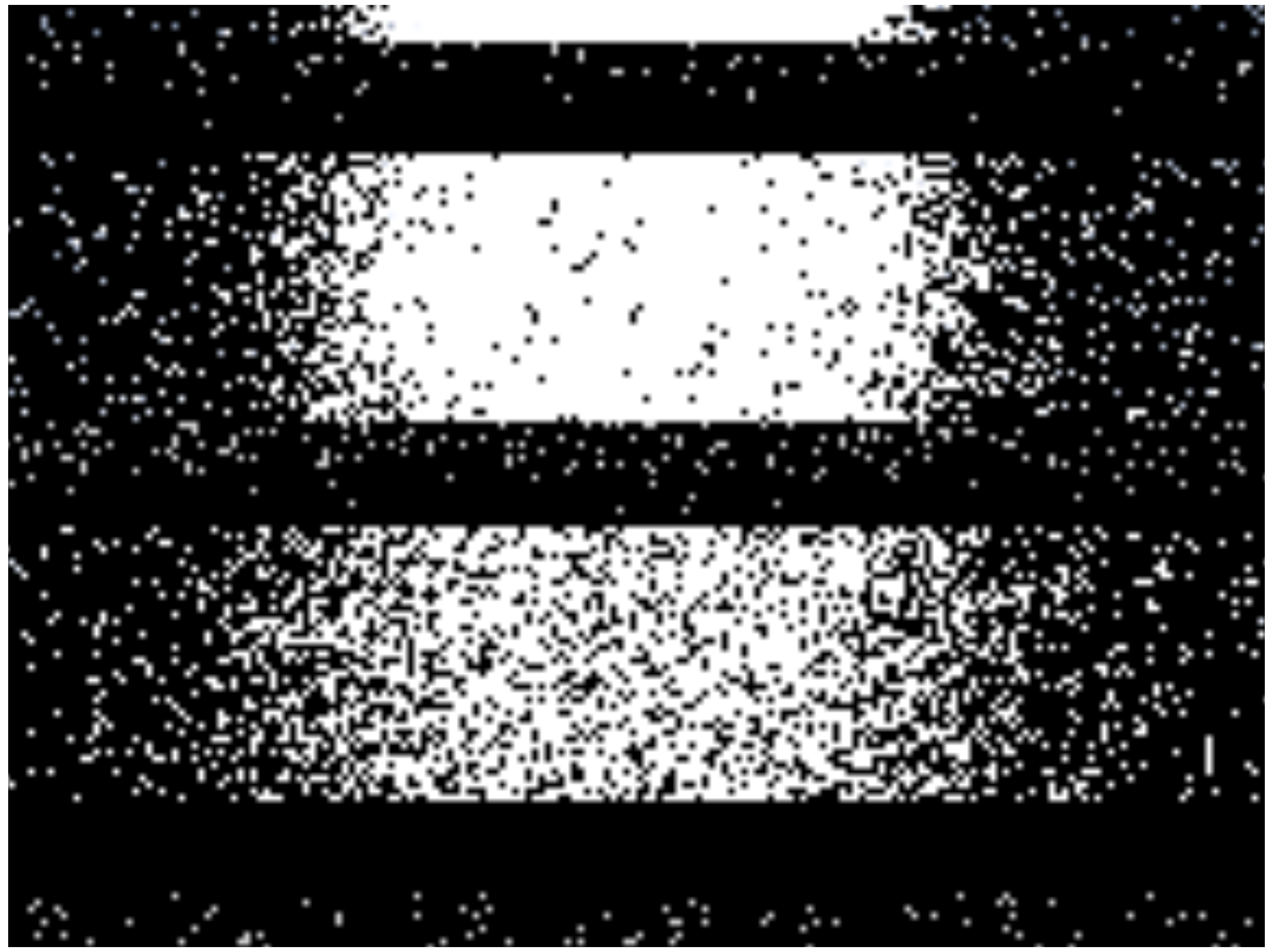
BSDF sampling is better in some regions

Variance reduction: Importance sampling



Reference image

N = 1024 spp



BSDF importance sampling

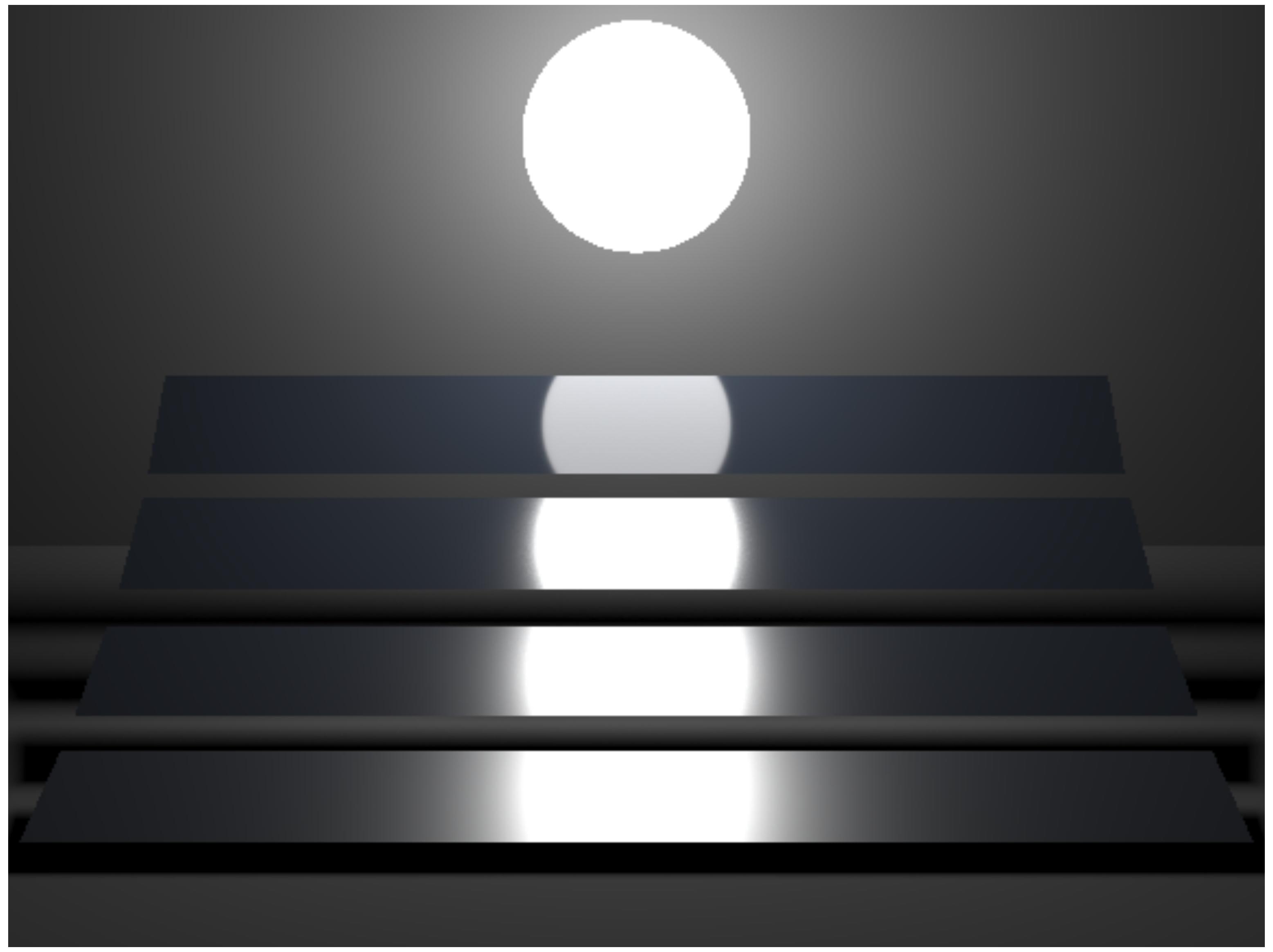
N = 4 spp

Light importance sampling

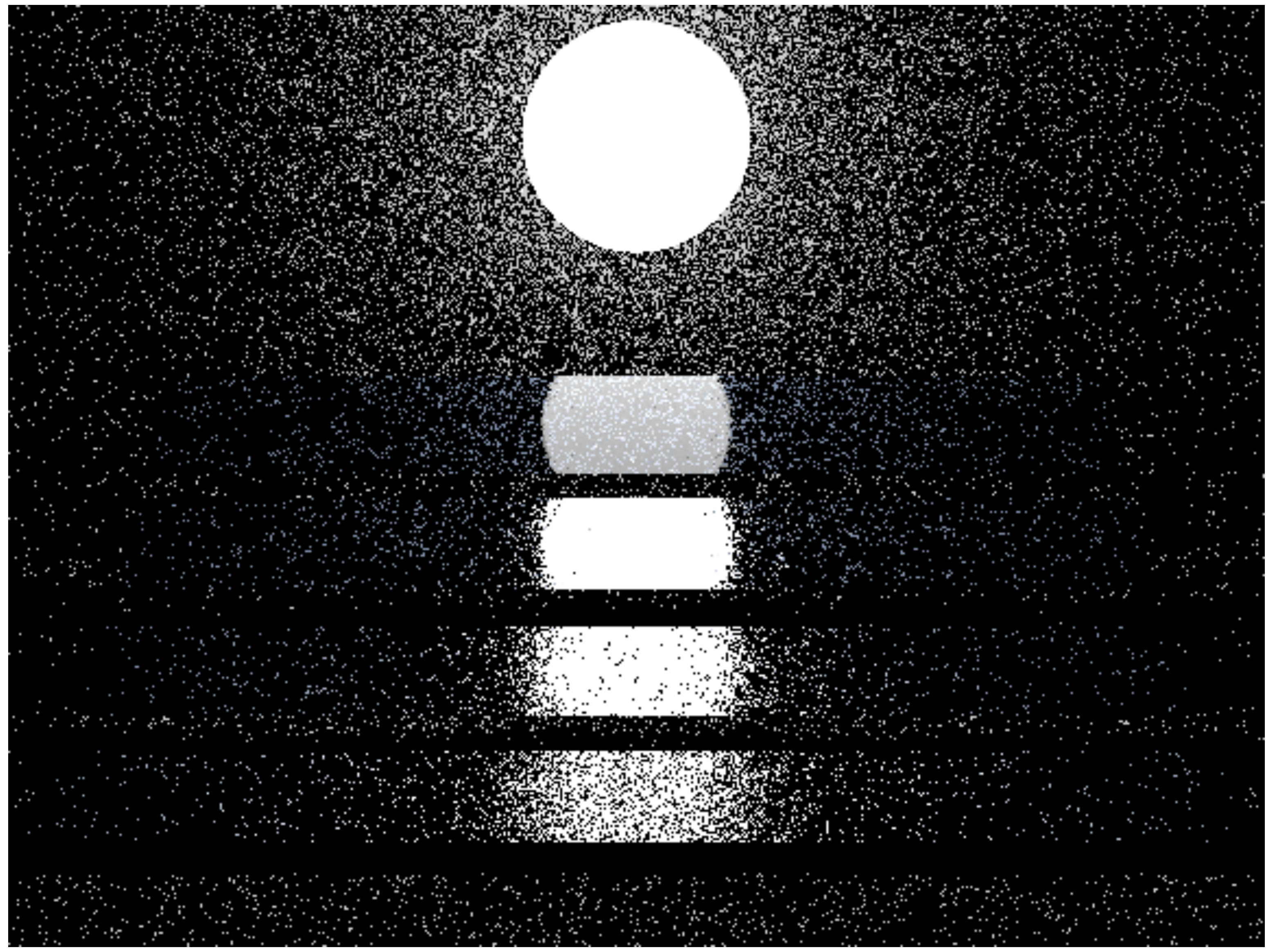
N = 4 spp

Light sampling is better in other regions

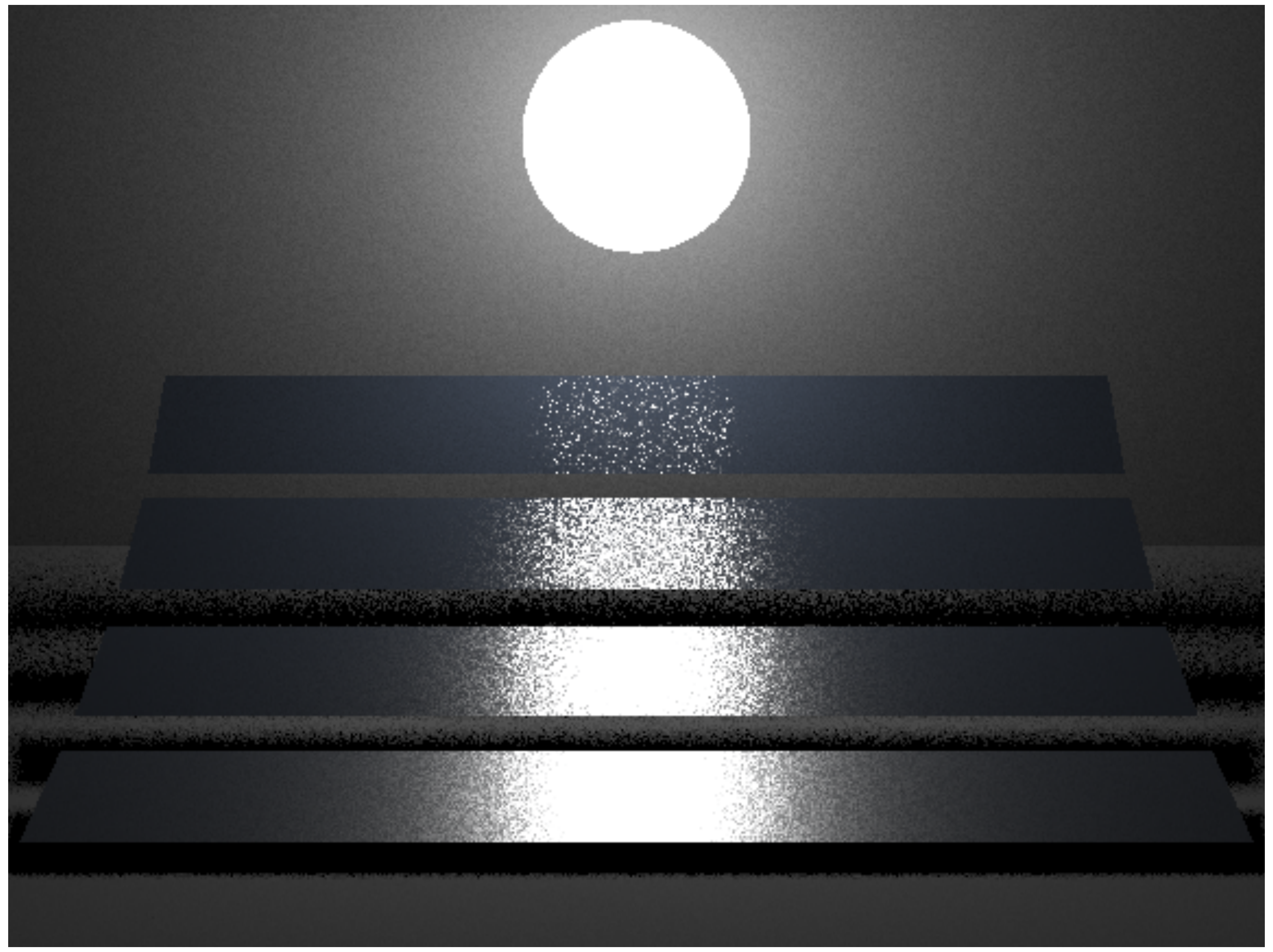
Variance reduction: Importance sampling



Reference image



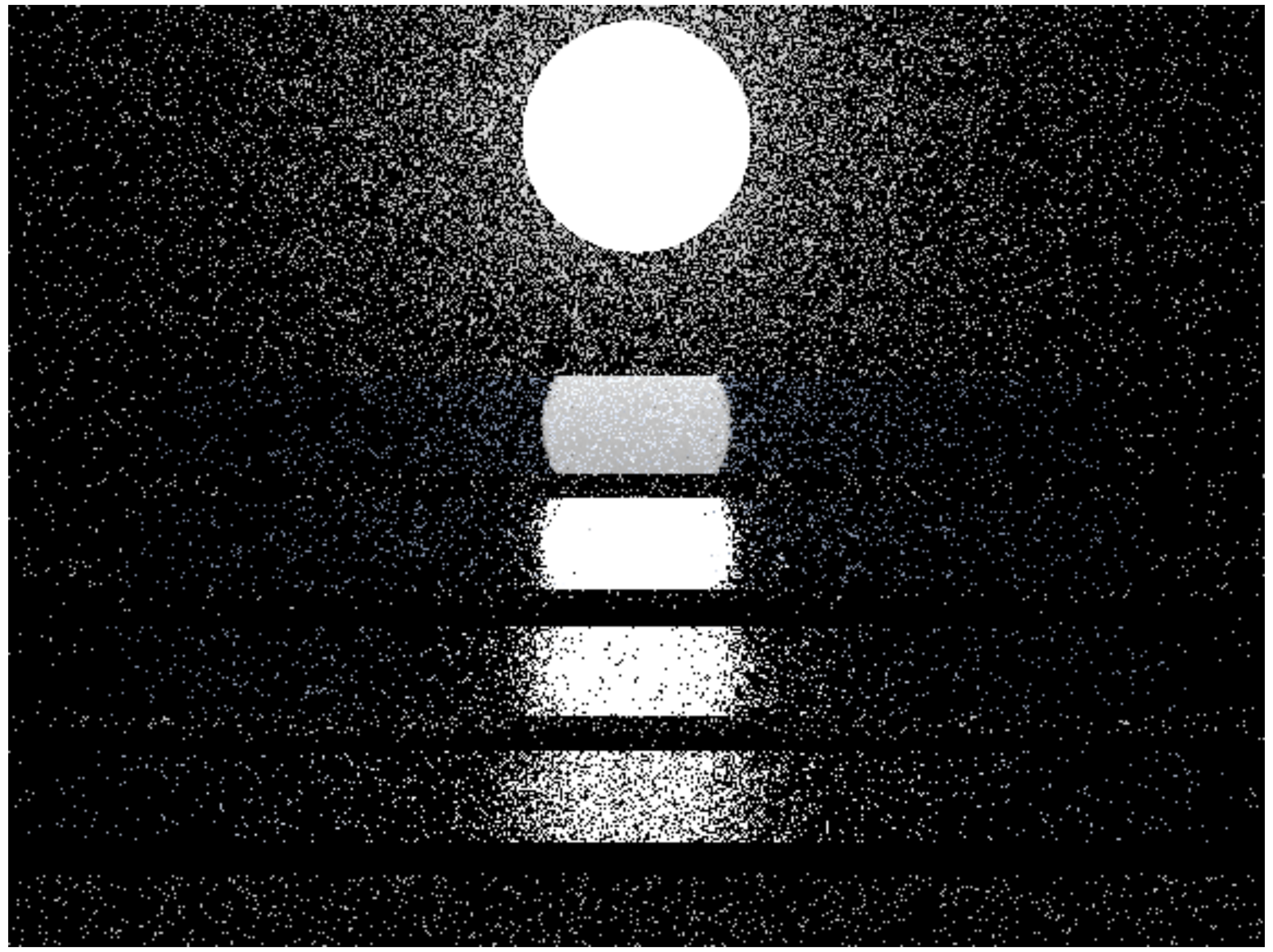
BSDF importance sampling



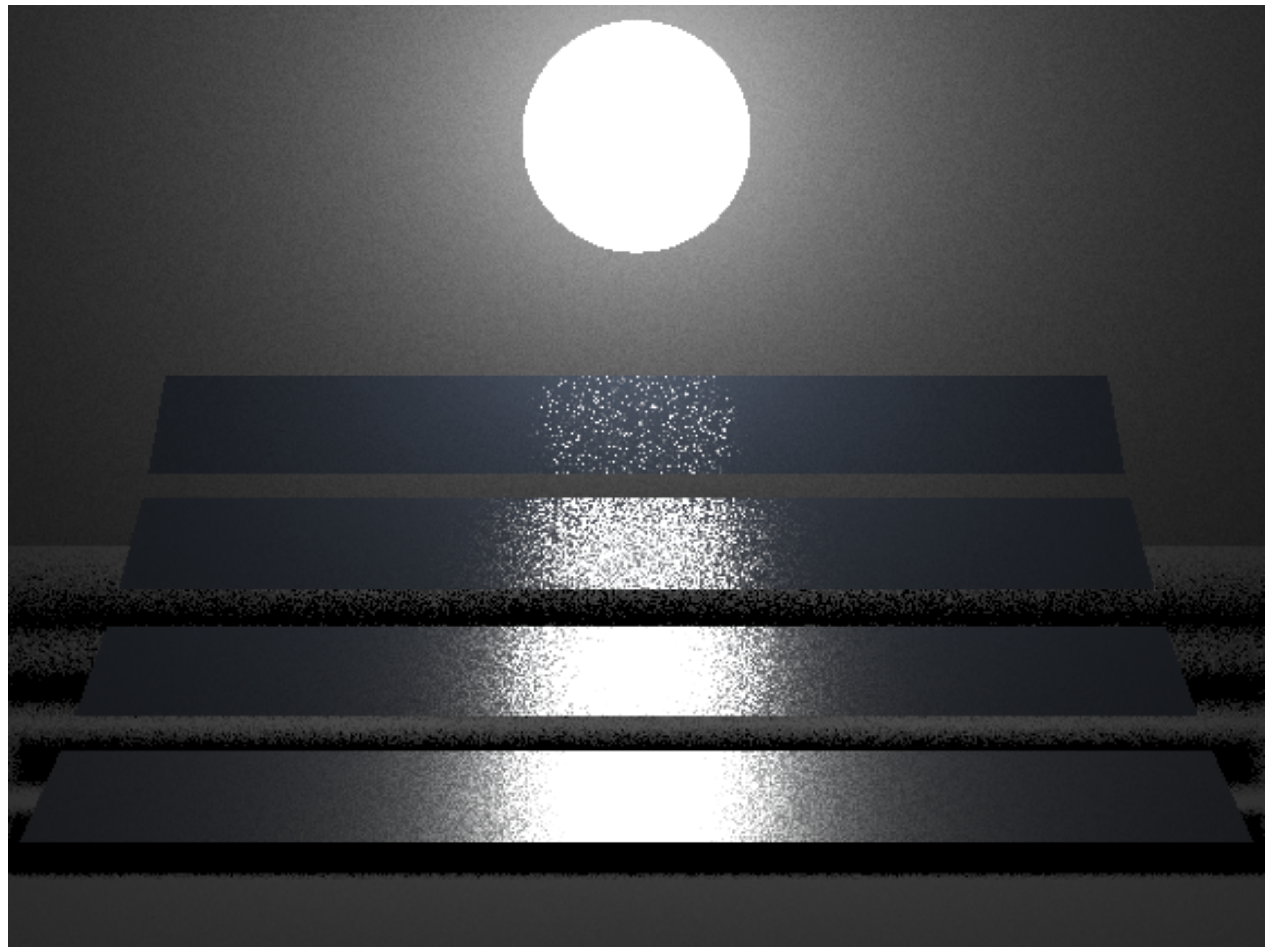
Light importance sampling

Can we combine the benefits of different PDFs ? **Yes!**

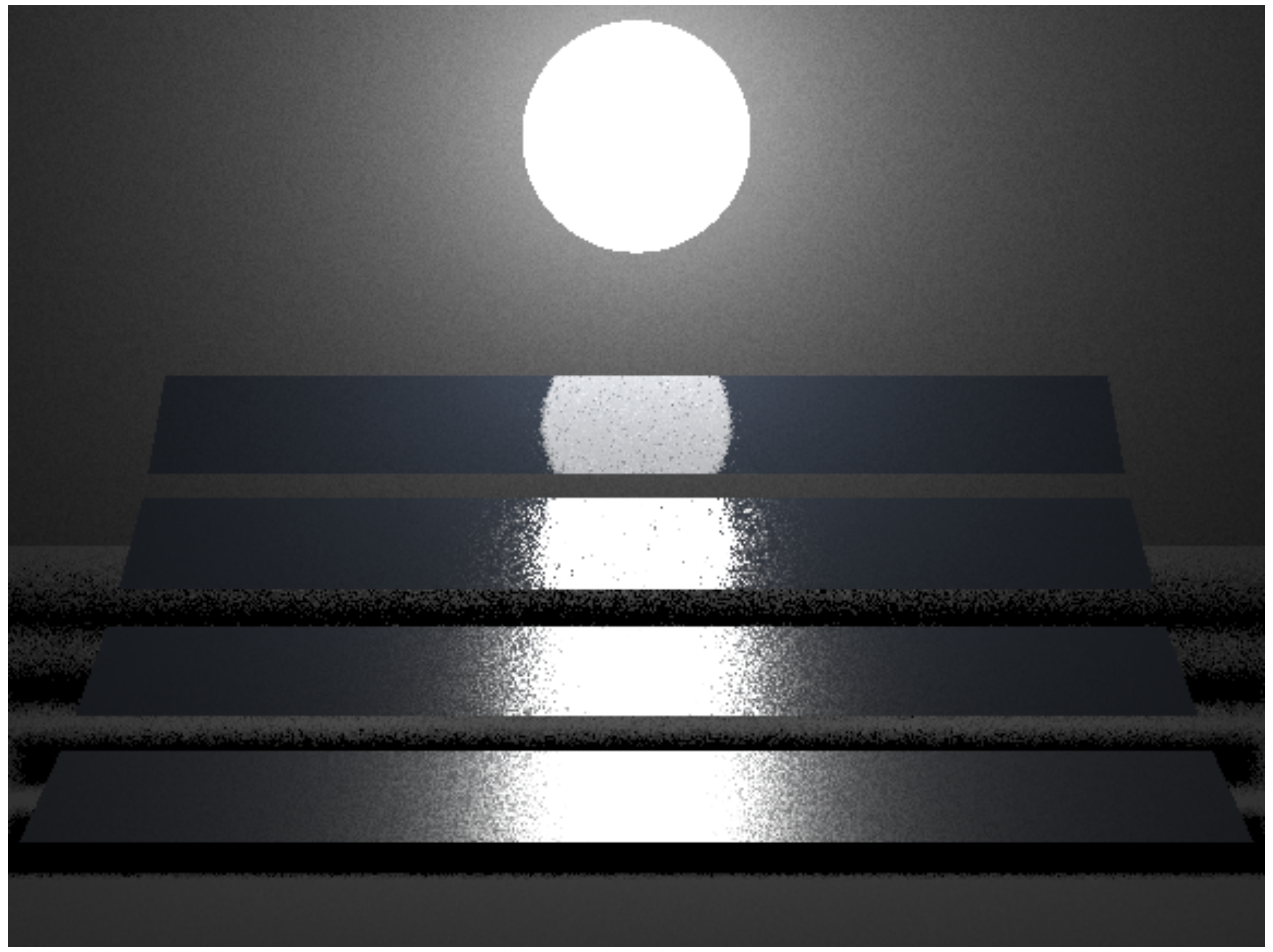
Variance reduction: Importance sampling



BSDF importance sampling



Light importance sampling



Multiple Importance Sampling

Can we combine the benefits of different PDFs ? **Yes!**

To be continued ... in the next lecture.